# Studies on the protein biology of the plant pathogen, *Candidatus* Liberibacter solanacearum

---

A thesis submitted in partial fulfilment of the

requirements for the degree of

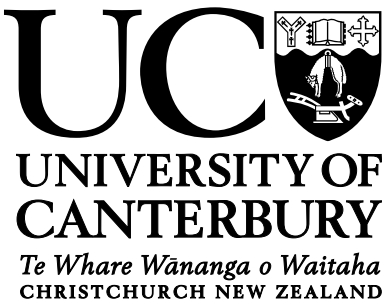
Doctor of Philosophy in Biochemistry

in the School of Biological Sciences

by Jenna Marion Gilkes

University of Canterbury

---




2019

# Abstract

*'Candidatus* Liberibacter solanacearum' (*C*Lso) is an unculturable α-proteobacterium that is the putative causal agent of zebra chip disease of potato. This disease is a major problem in potato growing areas in the United States and New Zealand, as it affects crop growth and yield, resulting in the loss of millions of dollars to the potato industries. Zebra chip disease is characterised by browning and necrotic flecking of vascular and internal tissue, which when fried, present as dark stripes and streaks within the chip rendering them commercially unacceptable. A major limitation to understanding the biology and pathology of *Liberibacter* species is the inability to culture the bacteria *in vitro.* Thus, in order in to further understand the biochemistry of the organism, a metagenomic approach was used to obtain the genome sequence of *C*Lso (Lin et al. 2011).

The potato-infecting clades of this bacterium are vectored by *Bactericera cockerelli*, (psyllid). Current disease management strategies target the psyllid: as the pathogen is transmitted relativity quickly, these strategies are limited in control of the disease. Thus, new management strategies that target the bacterial pathogen are required. However, the development of effective treatments for *C*Lso has been hampered by the difficulty in functionally characterizing the proteins of this organism, largely because they are not easily expressed and purified in standard expression systems.

A major aim of this thesis is to develop a method for the isolation of soluble and functional *C*Lso proteins. *C*Lso has a reduced genome and its proteins are predicted to be prone to instability and aggregation. Small genomes generally lack genes for DNA repair and have a bias towards A + T residues (Ochamn and Kelkar 2013). This, in turn, translates to a mutational bias towards hydrophobic residues at the protein level, increasing the likelihood of protein aggregation (Ochman and Kelkar 2013). Among intracellular-dwelling bacteria, chaperone proteins are conserved and overexpressed to buffer against problems in protein folding. An approach to express and purify soluble and functional *C*Lso proteins in *Escherichia coli* was established *via* the use of chaperone co-expression to mimic *in vivo* cellular conditions. This approach was utilised for the expression of *C*Lso adenyltransferase/IMP cyclohydrolase.

Characterisation determined the protein was folded and not functional. Solution structural analysis found that *C*Lso adenyltransferase/IMP cyclohydrolase exists as a monomer in solution, which contrasts to the dimer observed in other bacterial homologues. Dimerisation is required for activity of the enzyme and the oligomeric state of the enzyme was determined to account for the lack of activity in *C*Lso adenyltransferase/IMP cyclohydrolase.

Chaperone co-expression resulted in the isolation of an additional five *C*Lso proteins: serralysin (virulence factor), dehydrorhamnose epimerase, dehydrorhamnose reductase (L-rhamnose biosynthesis), pyruvate kinase (glycolysis) and dihydrodipicolinate synthase (lysine biosynthesis), which are hypothesised to be required for the organisms' survival since gene expression studies show high bacterial expression when in infected psyllids and *in planta.* Characterisation of purified dihydrodipicolinate synthase and pyruvate kinase showed that both enzymes were active and have the canonical tetrameric oligomeric structure in solution, consistent with other bacterial homologues.

The establishment of a robust method for the expression of *C*Lso proteins allows for the structural and kinetic characterisation of dihydrodipicolinate synthase, a key enzyme as it catalyses the committed step and is a potential antimicrobial target. The structure of the *C*Lso dihydrodipicolinate synthase was solved in complex with the substrates: pyruvate, pyruvate and succinic semi-aldehyde (an analogue of (*S*)-aspartate semialdehyde) and with the allosteric inhibitor, lysine. Structural analysis showed that there was little difference in the *C*Lso dihydrodipicolinate synthase compared with the *Agrobacterium tumefaciens* orthologue. Kinetic analysis of the enzyme reports constants that suggest a high binding affinity for substrates. It was proposed that *C*Lso dihydrodipicolinate synthase has evolved a high binding affinity for its substrates in response to nutrient limitation, which could occur as a result of the intracellular lifestyle of the organism.

Overall, this work provides new data that enriches our understanding of *C*Lso biology and the relationship of the organism with the psyllid host and provides a method for future characterisation of *C*Lso proteins, enabling new research and drug discovery programmes to study and manage the pathogenicity of the organism.

# Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisors: Professor Renwick Dobson, Dr. Grant Smith, Dr. Campbell Sheen and Dr. Rebekah Frampton, without you, this thesis would not have been possible. Thank you Grant and Rebekah, for you never ending support and guidance. Your insight and sharp eye for detail have been hugely appreciated.

To Ren Dobson, thank you for your unwavering support and encouragement throughout the past four years. I am so grateful to have had the opportunity to work with you. Thank you for your continued optimism, even when I did not see a way forward.

To Campbell Sheen, thank you for your support, encouragement and pep talks and for being a voice of reason and sanity. This work would not have been possible without your help and guidance.

To the Dobson Lab members, I feel privileged to have worked with such an amazing and supportive group of people. A huge thank you to Rachel, Jen, Chris, James and Michael Currie for always sparing the time to offer advice and help in so many ways. Thank you for your support and friendship over the years. To Amanda, Michael Love and David thank you for your endless support and encouragement.

Thank you to the Callaghan Innovation and the Biomolecular Interaction Centre for providing financial assistance over the past four years.

Thank you to my family, for your love, support and encouragement. I owe so much to you all. To my mother, thank you for your patience when I tried to explain my thesis to you and always knowing what to say when times were trying.

Lastly, lots of love and thanks to Hani. You have for being there for me at the best and worst of times. Thank you for your unwavering support and endless laughs. You made the last couple of years so much easier and more enjoyable.

# Table of Contents

# Abbreviations

| | |
|---|---|
| Å | angstrom |
| ° | degree |
| °C | degree celcius |
| Δ | delta |
| 2–ΔΔCt | double delta cycle threshold |
| µg | microgram |
| µL | microlitre |
| µM | micromolar |
| % | percent |
| σ | sigma |
| θ | theta |
| (β/α) | 8-barrel triosephosphate isomerase barrel |
| $A_{280}$ | absorbance at 280 nm |
| ADP | adenosine diphosphate |
| AIC | Akaike information criterion |
| AICAR | 5-amino-4-imidazolecarboxamide ribonucleotide |
| AICARFT | 5-amino-4-imidazolecarboxamide ribonucleotide formyl transferase |
| ATIC | adenyltransferase/IMP cyclohydrolase |
| ATP | adenosine triphosphate |
| AUC | analytical ultracentrifugation |
| β | beta |
| BSA | bovine serum albumin |
| B-factor | estimation of atomic mobility |
| BLAST | basic local alignment search tool |
| *C*Lso | *Candidatus* Liberibacter solanacearum |
| CD | circular dichroism |
| cDNA | complementary deoxyribonucleic acid |
| cm | centimetre |
| c(*M*) | continuous mass distribution |
| c(*s*) | continuous sedimentation coefficient distribution |
| CV | column volume |
| Da | dalton |
| DAP | diaminopimelic acid pathway |
| $ddH_20$ | double distilled water |
| DDM | n-dodecyl-β-D-maltoside |
| DTT | dithiothreitol |

| | |
|---|---|
| DHDPS | dihydrodipicolinate synthase |
| DHDPR | dihydrodipicolinate reductase |
| DHRE | dehydrorhamnose epimerase |
| DHRR | dehydrorhamnose reductase |
| Dmax | maximum particle diameter |
| DNA | deoxyribonucleic acid |
| DSF | differential scanning fluorimetry |
| EDTA | ethylenediaminetetraacetic acid |
| $f/f_0$ | frictional ratio |
| FAICAR | 5-formylaminoimidazole-4-carboxamide ribonucleotide |
| g | gram |
| GFP | green fluorescent protein |
| h | hour |
| HEPES | *N*-2-hydroxyethylpiperazine-*N*'-2-ethane sulfonic acid |
| His-tag | Histidine tag |
| $His_6$ | hexa-histidine tag |
| HRV3C | human rhinovirus 3C protease |
| $I_0$ | scatter intensity at the origin |
| IMAC | immobilised metal affinity chromatography |
| IMPCH | inosine monophosphate cyclohydrolase |
| IPTG | isopropyl β-D-1-thiogalactopyranoside |
| Kbp | kilo base pairs |
| $k_{cat}$ | catalytic turnover number |
| $K_m$ | Michaelis-Menten constant (rate of enzyme at half Vmax) |
| kDa | kilodalton |
| $K_i$ | inhibition constant |
| L | litre |
| LB | Luria Bertani |
| LDH | lactate dehydrogenase |
| M | molar |
| Mbp | mega base pairs |
| MES | 2-(N-morpholino)ethanesulfonic acid |
| mg | milligram |
| min | minute |
| mL | millilitre |
| mm | millimetre |
| mM | millimolar |
| mol | moles |
| MRE | mean residue ellipticity |
| MW | molecular weight |
| NADH | nicotinamide adenine dinucleotide, reduced form |

| | |
|---|---|
| NADPH | nicotinamide adenine dinucleotide, reduced form |
| nm | nanometre |
| $OD_{600}$ | optical density at 600 nm |
| *P(r)* | real space distance distribution function |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| PK | pyruvate kinase |
| qRT | quantitative reverse transcription |
| r | radius |
| $R_{factor}$ | residual factor |
| $R_{free}$ | free Rfactor |
| $R_g$ | radius of gyration |
| rmsd | root-mean-square deviation |
| RNA | ribonucleic acid |
| rpm | revolutions per minute |
| S | sedimentation coefficient, Svedberg |
| s | second |
| s20,w | standardised sedimentation coefficient relative to water at 20 °C |
| [S] | substrate concentration |
| (*S*)-ASA | S-aspartate semialdehyde |
| SAXS | small angle X-ray scattering |
| SDS | sodium dodecyl sulfate |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SEC | size exclusion chromatography |
| SOC | catabolite repression super optimal broth |
| T state | tense state |
| TAE | tris-acetate-ethylenediaminetetraacetic acid |
| TCEP | tris(2-carboxyethyl)phosphine |
| TF | trigger factor protein |
| $T_M$ | melting temperature |
| Tris | 2-amino-2-hydroxymethyl-propane-1,3-diol |
| UDP | uridine diphosphate UV ultraviolet |
| $V_0$ | initial velocity |
| $V_M$ ($Å^3$ $Da^{-1}$) | wilson B value |
| $V_{max}$ | maximum rate of an enzyme |
| v/v | volume to volume |
| w/v | weight to volume |

# Chapter One

# Introduction

The following is a general introduction to the field of *Candidatus* Liberibacter solanacearum biology and pathogenicity. Subsequent chapters include more detailed, specific introductions to the chapter topics, as these are intended to be independent bodies of work. A version of this introduction was published as a review article:

Gilkes, J.M., Frampton, R.A., Smith, G.R., and Dobson, R.C.J. 2018. Potential pathogenicity determinants in the genome of '*Candidatus*' Liberibacter solanacearum, the causal agent of Zebra Chip disease of potato. Australasian Plant Pathology 47(2)119-134.

## 1.1 Zebra chip disease

Zebra Chip disease is a recently identified, economically important disease of potato (*Solanum tuberosum L*.). Zebra chip disease was first identified in 2000 in commercial potato fields in Pearsall and the Lower Rio Grande Valley in Texas, USA (Secor and Rivera-Varas 2004) and subsequently has been observed in Nebraska, Colorado, Kansas, New Mexico, Arizona, Nevada and California (Munyaneza et al. 2007). Symptoms of the bacteria were observed in potato tubers harvested from South Auckland, New Zealand, in 2008 (Liefting et al. 2008b). Later that year the first visual symptoms of zebra chip disease in processed potatoes were recorded in Pukekohe, NZ. Although the vector, *Bactericera cockerelli* Šulc (tomato potato psyllid), spread to the South island, symptoms of zebra chip disease were only visible during the 2013 – 2014 growing season (Jorgensen et al. 2013). Zebra chip has had a serious economic impact on the commercial potato industry in New Zealand (Liefting et al. 2008a; Liefting et al. 2008b; Teulon et al. 2009; Ogden 2011). Recently, the vector for zebra chip disease, the tomato potato psyllid, was identified in Perth, Australia, posing a new threat to the country's biosecurity and potato production (Fitzgerald 2017).

Zebra chip is characterised by the development of dark stripes and streaks when potato tubers are fried to produce potato chips (Figure 1.1). The matrix of dark and light regions of the chips renders them commercially unacceptable (Munyaneza et al. 2007; Secor et al. 2009; Crosslin et al. 2010; Miles et al. 2010; Buchman et al. 2011; Munyaneza 2012). The streaks usually coincide with the medullary rays of the tuber (Crosslin et al. 2010). Foliar symptoms of the disease include upward rolling of the leaflets, purple and yellow discoloration, proliferation of axillary buds, shortened buds, leaf scorch, swollen nodes, aerial tubers and eventual senescence (Munyaneza 2007; Secor et al. 2009; Crosslin et al. 2010; Munyaneza 2012).

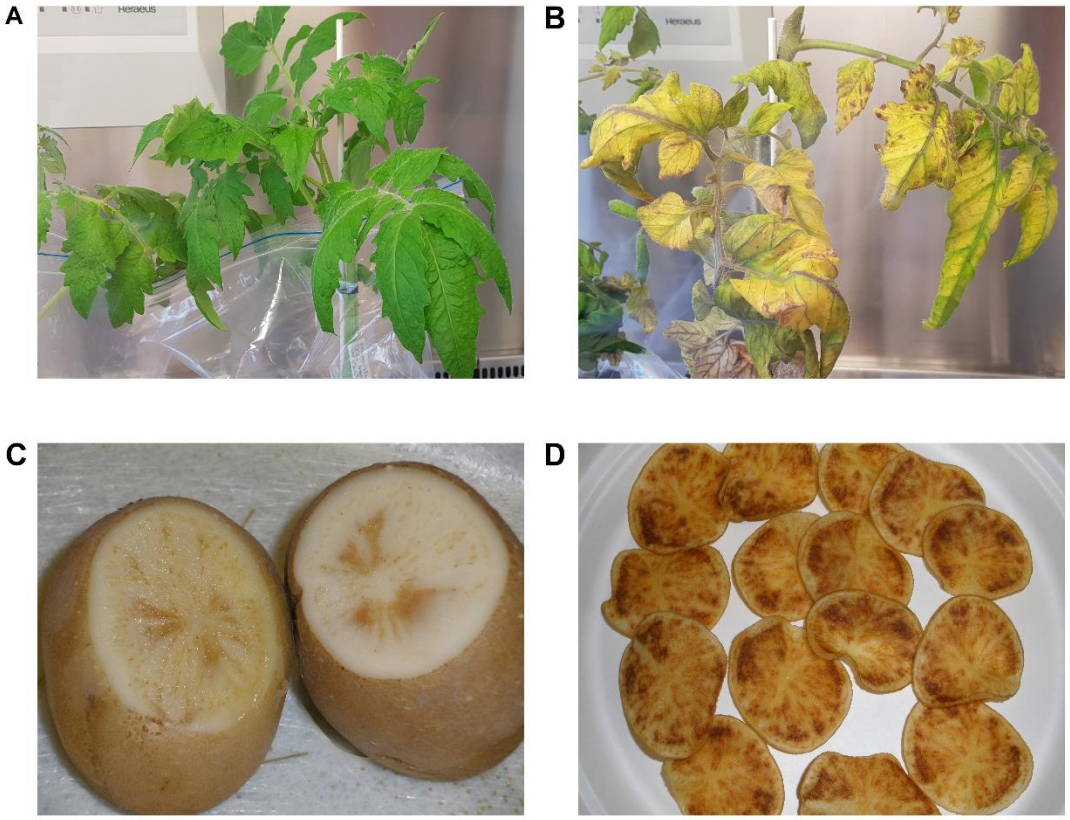



**Figure 1.1:** Zebra chip disease symptoms. **A** is an uninfected tomato plant whilst **B** shows a tomato plant with zebra chip disease foliar symptoms. **C** potato tubers showing zebra chip disease in raw potato tubers. **D** potato tuber after frying showing the dark stripes and streaks characteristic of zebra chip disease. (Images C and D provided by the New Zealand Institute for Plant and Food Research).

## 1.2 Disease management strategies

The primary management strategy for zebra chip disease is the application of insecticides to control the psyllid. Although the application of these insecticides and potential new bio-rational insecticides are effective at reducing tomato potato psyllid populations, transmission of *Ca*. L. solanacearum (*C*Lso) to the psyllid through feeding on plant phloem occurs within a few hours (Buchman et al. 2011; 2012), thus limiting the effectiveness of this strategy even if applied immediately upon detection of insects in the field.

Broad spectrum insecticides have potential deleterious effects on non-target insects, with the potential to disrupt biocontrol or pollination services. The introduction of insect pathogens as a component of integrated pest management offers long-term sustainable psyllid control with limited environmental off-target effects (Lacey et al. 2011). Entomopathogenic fungi are

the only pathogens used for the control of Hemiptera because their piercing and sucking mouthparts do not allow infection by pathogens that require ingestion in order to infect the host (Lacey et al. 2011). Lacey et al. (2011) demonstrate effective control of adult and nymph psyllids with the use of *Metarhizium anisopliae* and *Isaria fumosorosea*, with reductions of the tomato potato psyllid eggs and nymphs of 45% and 59% respectively.

Within New Zealand, two naturalised coccinellids, *Cryptolaemus montrouzieri* and *Cleobora mellyi* were investigated as biological control agents. O'Connell et al. (2012) found that both *C. monotrouzieri* and *C. mellyi* were capable of ingesting 30 and 100 psyllids, respectively, over a 24 hour period, indicating that a new predator-prey association may exist. In June 2016, the New Zealand Environmental Protection Agency approved the release of the bio-control agent *Tamarixia trioaze* into potato, outdoor tomato, and tamarillo production areas. *Tamarixia trioaze* is a parasitoid wasp and is found in the USA and Mexico as a natural predator of the tomato potato psyllid. This provides a potential cost-effective solution for potato, tomatoes and capsicum growers in New Zealand.

Zebra Chip resistant or tolerant varieties of potato varieties could also provide an efficient and sustainable means of zebra chip disease management (Munyaneza et al. 2011). Commercially available potato varieties tested in 2009 and 2010 were highly susceptible to the pathogen (Munyaneza et al. 2011) with all plants developing severe zebra chip foliar and tuber symptoms. Research is in progress to produce resistant potato cultivars and to identify potato varieties and advanced breeding lines with resistance to the bacterium and/or the psyllid (Munyaneza et al. 2011; Butler and Trumble 2012).

The emergence and impact of zebra chip disease has highlighted the need for effective bacterial management strategies. Many of the disease management plans are targeted at the psyllid vector; few strategies target *C*Lso. The publication of the first *C*Lso genome in 2011 (Lin et al. 2011) offered the potential to propose new control strategies and identify novel targets for antibacterial drug design.

## 1.3 *Candidatus* Liberibacter solanacearum

*Candidatus* Liberibacter solanacearum (Liefting et al. 2009) was first described in 2008 in New Zealand in potato, capsicum, tomato (Liefting et al. 2008a; Liefting et al. 2009), cape gooseberry and tamarillo (Liefting et al. 2008b) after investigation of the aetiology of potential disease symptoms on glass house-grown tomatoes (*Solanum lycopersicum*) and capsicum (*Capsicum annuum*) in New Zealand. It is considered to be synonymous with *Candidatus* Liberibacter psyllaurous which was described by Hansen et al. (2008) from the tomato potato psyllid, tomato and potato in the USA. *C*Lso has also been associated with disease symptoms on the apiaceous plants carrot and celery (Munyaneza et al. 2010a; Teresani et al. 2014). *C*Lso is vectored by three Triozidae psyllids: the tomato potato psyllid, *B. cockerelli* (Munyaneza et al. 2008) in the United States, New Zealand and Norfolk Island (Australia), the carrot psyllid, *Trioza apicalis* in Finland (Munyaneza et al. 2010b) and *Bactericera trigonica* in the Canary Islands (Alfaro-Fernandez et al. 2012).

Infestations of the psyllid are historically rare, with the vector reported in central North America during the summer months of 1928 (Richards 1928). Sporadic outbreaks occurred in California during the 1930s and 1940s (Pletsch 1947), with an outbreak of *B. cockerelli* and psyllid yellows occurring in 1938 (Jensen 1939). In the years after the outbreak, *B. cockerelli* was managed exclusively by insecticides (Pletsch 1947; Wallis 1955).

The tomato potato psyllid was first discovered in Auckland, NZ, in 2006 (Gill 2006) and since then, it has spread throughout New Zealand to become a major economic pest. The psyllid was first reported in mainland Australia in February 2017 in metropolitan Perth (Fitzgerald 2017). The tomato potato psyllid is considered to be the main vector of *C*Lso (Liefting et al. 2008a; Liefting et al. 2008b; Hansen et al. 2008) (Figure 1.2A). The psyllid acquires the bacterium by feeding on infected solanaceous plants. To be persistently transmitted, *C*Lso must pass through the alimentary canal wall then move through the haemolymph to the salivary glands where the bacteria can then be transmitted to new host plants during psyllid feeding (Haapalainen 2014; Prager et al. 2015). Sandanayaka et al. (2014) found that the plant access period for a single psyllid adult to transmit *C*Lso was two hours whilst twenty adults were capable of inoculating a potato plant within an hour to result in Zebra Chip disease

symptoms (Buchman et al. 2011). The concentration of *C*Lso within the alimentary canal of the psyllid remains constant over time. However, two weeks after pathogen acquisition the concentration of bacteria peaks in the salivary glands. Furthermore, it is only after this two week period that the psyllids are capable of transmitting *C*Lso to non-infected host plants (Sengoda et al. 2013; Sengoda et al. 2014). Adult and fifth instar nymph psyllids carry the highest titre of *C*Lso (Casteel et al. 2012). Despite young nymphs having a lower bacterial titre, Casteel et al. (2012) found they were still capable of transmitting the bacteria.

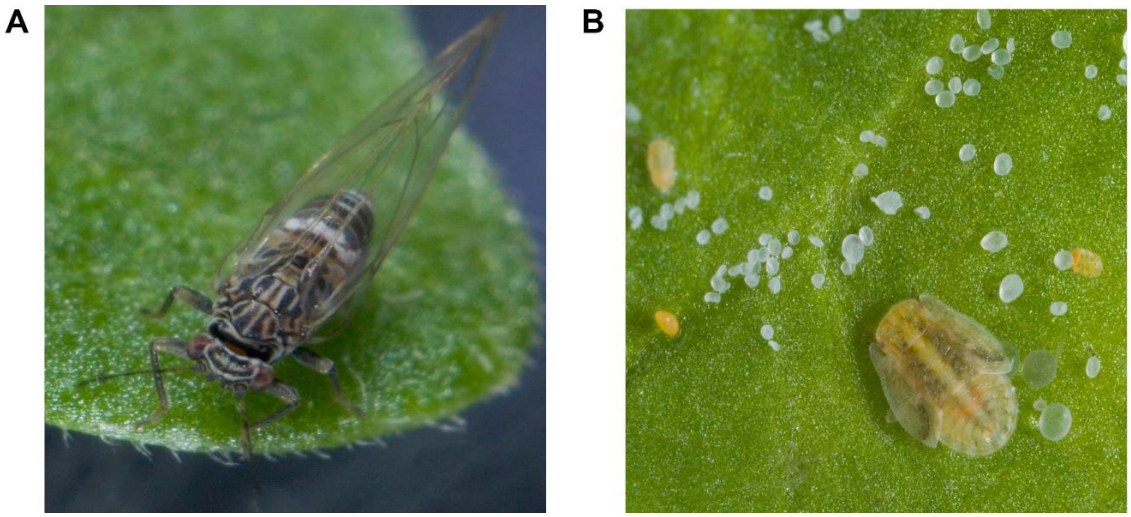



**Figure 1.2:** The *C*Lso psyllid vector, *Bactericera cockerelli*. **A** is the adult psyllid and **B** is the *B. cockerelli* nymph. (Images provided by the New Zealand Institute for Plant and Food Research).

## 1.4 Introduction to *Candidatus* Liberibacter genus

*Candidatus* Liberibacter species are recognized as the causal agents of a number of destructive plant diseases around the world, including Zebra Chip disease (Munyaneza et al. 2007; Liefting et al. 2008a; Liefting et al. 2008b; Abad et al. 2008; Liefting et al. 2009) and huanglongbing (Coletta-Filho et al. 2004), which currently pose a serious threat to the potato, tomato, carrot and citrus industries. The genus *Liberibacter* currently consists of seven unculturable species with *Candidatus* status (L. solanacearum, L. asiaticus, L. americanus, L. africanus, L. carribeanus and L. europaeus) and one culturable full status species, *L. crescens* (Table 1.1). In 2017, sequences with 99% identity to *Ca.* Liberibacter 16S rRNA sequences were found in the Australian native psyllid *Acizzia solanicola* by Morris et al. (2017) with a proposed name of *Candidatus* Liberibacter brunswickensis (Table 1.1).

**Table 1.1:** Currently described *Liberibacter* species and associated vectors

| Species | Clade/Region | Vectors | Plant host family |
|---|---|---|---|
| *Ca*. L. solanacearum | A-Americas and New Zealand | *Bactericera cockerelli* (Triozidae) | Solanaceae (various species) |
| | B- Americas | | |
| | C-Scandinavia | *Trioza apicalis* (Triozidae) | Apiaceae (Carrot) |
| | D- Europe | *Bactericera trigonica* (Triozidae) | |
| | E- Europe | Unknown (possibly *Bactericera trigonica*) | Apiaceae (Celery) |
| *Ca*. L. asiaticus | | *Diaphorina citri* (psyllidae)<br>*Trioza erytreae* (Triozidae)<br>*Cacopsylla citrisuga* (Psyllidae) | Rutaceae (various species) |
| *Ca*. L. americanus | | *Diaphorina citri* (Psyllidae) | |
| *Ca*. L. africanus | | *Trioza erytreae* (Triozidae)<br>*Diaphorina citri* (Psyllidae) | |
| *Ca*. L. caribbeanus | | *Diaphorina citri* (Psyllidae) | |
| *Ca*. L. europaeus | | *Arytainilla spartiophila* (Psyllidae) | None described |
| *Liberibacter crescens* | | None described | None described |
| *Ca*. L. brunswickensis | | *Acizzia solanicola* (Psyllidae) | Solanaceae (Various species) |

*Ca.* L. asiaticus was the first species in this genus to be identified. Lafleche and Bove (1970) observed a prokaryote in the phloem of leaves from sweet orange plants that were showing symptoms of huanglongbing. Initially, this prokaryote was believed to be a mycoplasma-like organism, however transmission electron microscopy examination of samples revealed a 25 nm thick envelope that distinguished this organism from the mycoplasmas (Lafleche and Bove 1970). The organism was phloem limited and it is not culturable. Garnier et al. (1984) established that the bacterium-like organism associated with citrus greening was a Gram-negative bacterium, and Jagoueix et al. (1994) placed it in the α subdivision of the Proteobacteria based on analysis of 16S rRNA sequence.

Raddadi et al. (2011) described a novel *Liberibacter* species in Europe present in the pear psyllid pest *Cacopsylla pyri* and designated it *Candidatus* L. europaeus. This species was considered to be an endophyte, rather than a pathogen, of pear as it was present in high titre of over $10^8$ 16S rRNA copies $g^{-1}$ in pear plant tissue but infected plants did not display any disease symptoms. In New Zealand, *Ca*. L. europaeus was detected in Scotch broom (*Cytisus scoparius*) infested by the broom psyllid, *Arytainilla spartiophila* (Thompson et al. 2013). The plants displayed symptoms such as stunted growth of shoots, leaf tip chlorosis and leaf dwarfing (Thompson et al. 2013).

## 1.5 The *Liberibacter* genome

Five haplotypes of *C*Lso which have been described based on SNPs and insertion-deletions (indels) in the rRNA region (Table 1.2) (Nelson et al. 2011; Teresani et al. 2014). The haplotype A currently denoted NZ1 (JMTK00000000)and HenneA (JQIG00000000) (Clade 1) (Thompson et al. 2015) and B (ZC-1) (NC_014774) (Clade 2) (Lin et al. 2011) are associated with solanaceous crops and the tomato potato psyllid in Central and North America whereas only haplotype A occurs in New Zealand (Nelson et al. 2011; Thompson et al. 2015). Haplotype C is associated with carrot and *T. apicalis* in Scandinavia and two genomic sequences of haplotype C have been determined FIN114 (LWEB00000000) and FIN111 (LVWB01000000) (Wang et al. 2017a). Haplotype D is found in carrot and celery in Europe and is vectored by *B. trigonica* (Teresani et al. 2014). There are two other draft genome sequences of *C*Lso, R1 (JNVH00000000) and RSTM (NZ_LLVZ01000000) (Wu et al. 2015; Wang et al. 2017a),

however, the sequences are highly fragmented which limits their use in genomic analysis. It is currently unknown which psyllid vector(s) haplotype E is associated with. Currently, there are seven *C*Lso genomes (Table 1.3) available in Genbank with the number of scaffolds ranging from one to 99. Genomic comparisons of the three haplotypes with sequenced genomes could help identify haplotype-specific genes that maybe involved in various host and psyllid interactions (Wang et al. 2017a).

**Table 1.2:** Genetic variation description systems for *Candidatus* Liberibacter solanacearum

| Classification type | Name | | | | | | Reference |
|---|---|---|---|---|---|---|---|
| rRNA | Haplotype | A | B | C | D | E | Glynn et al (2012) |
| | Clade | 1 | 2 | | | | Wen et al (2009) |
| MLST[a] | ST | 2 | 1 | | | | Nelson et al (2011) Teresani et al (2014) |
| SSR[b] | Lineage | 2 | 1 | | | | Lin et al (2012) |
| | Type | 1 | 2 | | | | Lin and Gudmestad (2013) |

[a]MLST, Multi Locus Sequencing

[b]SSR, Simple Sequence Repeat

A major limitation to understanding the biology and pathology of *Liberibacter* species is the inability to culture the bacteria *in vitro*. Li et al. (2009) reported that some heavily infected adult psyllids could harbour up to $10^{10}$ *Ca*. L. asiaticus cells per insect. Thus, a metagenomic approach using multiple displacement amplification (technique using Phi29 DNA polymerase to amplify small amounts of DNA) and pyrosequencing (sequencing technique which contrasts to Sanger sequencing as it relies on the detection of pyrophosphate release and the generation of light on nucleotide incorporation, rather than chain termination with dideoxynucleotides) was used to obtain the full bacterial genome for *Ca*. L. asiaticus (Duan et al. 2009) and *C*Lso (Lin et al. 2011). Currently, there are eighteen available *Ca*. Liberibacter species genomes (Table 1.3). A common feature of the *Ca*. Liberibacter genomes is the low guanine-cytosine content (34.5% in *Ca*. L. africanus, 36.6% in *Ca*. L. asiaticus and 31% in *Ca*. L. americanus) compared to the genomes of other phylogenetically close species in the Rhizobiaceae (e.g. 59% GC for *Agrobacterium tumefaciens* NZ_CP011246.1). The other key feature of the genomes of *Ca*. Liberibacter species is the reduced size, ranging from 1.1 – 1.3 Mbp, which is significantly smaller than those from other Rhizobiaceae such as *A. tumefaciens* (NZ_CP011246.1) with a genome size of 5.47 Mbp (Li et al. 2011).

**Table 1.3:** The available genomes of the (*Candidatus*) Liberibacter species

| Species | ID | Reference | Size (Mbp) | GC% | Genes | Annotation |
|---|---|---|---|---|---|---|
| *Ca.* L. solanacearum | NZ1 | JMTK00000000 | 1.312 | 35.3 | 1189 | RAST |
| *Ca.* L. solanacearum | ZC1 | NC_014774 | 1.258 | 35.2 | 1148 | NCBI_PGAAP[a] |
| *Ca.* L. solanacearum | R1 | JNVH00000000 | 1.204 | 35.3 | 1086 | RAST |
| *Ca.* L. solanacearum | HenneA | JQIG00000000 | 1.211 | 34.9 | 1093 | RAST |
| *Ca.* L. solanacearum | RSTM | NZ_LLVZ01000000 | 1.287 | 35.1 | 1172 | RAST |
| *Ca*. L. solanacearum | TX2351 | MTIM00000000 | 1.255 | 36.5 | 1184 | RAST |
| *Ca*. L. solanacearum | FIN111 | LVQB01000000 | 1.200 | 34.9 | 1040 | NCBI_PGAAP |
| | FIN114 | LWEB00000000 | 1.240 | 35.2 | 1067 | NCBI_PGAAP |
| *Ca.* L. asiaticus | psy62 | NC_012985 | 1.227 | 36.5 | 1162 | GeneMarks<br>FGENESB |
| *Ca.* L. asiaticus | gxpsy | NC_020549 | 1.268 | 36.6 | 1191 | NCBI_PGAAP |
| *Ca.* L. asiaticus | A4 | JFGQ01000000 | 1.234 | 36.4 | 1110 | RAST[b] |
| *Ca.* L. asiaticus | HHCA | JMIL01000000 | 1.150 | 36.5 | 1166 | RAST |
| *Ca.* L. asiaticus | Ishi-1 | NZ_AP014595.1 | 1.191 | 36.3 | 1064 | MiGAP[c] |
| *Ca.* L. asiaticus | FL17 | NZ_JWHA0100000 | 1.227 | 36.4 | 1101 | RAST |
| *Ca.* L. asiaticus | YCPsy | NZ_LIIM01000002 | 1.234 | 36.5 | 1111 | RAST |
| *Ca.* L. asiaticus | SGCA5 | NZ_LMT01000015 | 1.201 | 36.4 | 1091 | RAST |
| *Ca.* L americanus | Sao Paulo | NC_022793 | 1.195 | 31.1 | 1023 | NCBI_PGAAP |
| *Ca.* L americanus | PW_SP | AOFG01000000 | 1.176 | 31.1 | 1001 | NCBI_PGAAP |
| *Ca.* L. africanus | PTSAPSY | NZ_CP004021 | 1.192 | 34.5 | 1084 | NCBI_PGAAP |
| *L. crescens* | BT-0 | NZ_CP010522 | 1.522 | 35.4 | 1368 | NCBI_PGAAP |

[a]NCBI_PGAAP, National Centre for Biotechnology Information Prokaryotic Genome Automatic Annotation Pipeline

[b]RAST, Rapid Annotations using Subsystems Technology (Aziz 2008)

[c]MiGAP, Microbial Genome Annotation Pipeline (Katoh et al. 2014)

### 1.5.1 Inconsistencies in genome-wide protein function annotations

The relatively small, AT-rich genomes of the *Ca*. Liberibacter species currently in Genbank contain genes identified by various methods: Prokaryotic Genome Annotation Pipeline (PGAP), Rapid Annotations using Subsystems Technology (RAST), Microbial Genome Annotation Pipeline (MiGAP), GeneMarkS and FGENESB. The majority of sequences deposited in public databases have been analysed only by computation prediction, which raises the issue of annotation accuracy and database quality (Bork and Bairoch 1996; Karp 1998; Schnoes et al. 2009). Brenner (1999) and Devos and Valencia (2001) predicted that 8% of molecular function annotations were incorrect. Although genes can be annotated through bioinformatics analysis, there is still a need to relate the biochemical and cellular function to the uncharacterized gene (Johnston et al. 2003). Protein characterization and structural analysis determined that the annotated *M. tuberculosis menG*, predicted to encode an S-adenosylmethionine dependent methyltransferase, had no structural similarity to the suggested annotation (Johnston et al. 2003). This highlights the importance of gene validation through protein characterization in order to relate annotated gene to protein function in the *C*Lso genome, which will be a focus within this thesis.

## 1.6 Niche consequences for organisms with reduced genomes

*Ca.* Liberibacter species have reduced genomes which has a number of consequences. A reduced or minimal genome is defined as the gene set required for life under nutrient rich and stress free conditions (Moran 2002; McCutcheon and Moran 2012). Genomes of less than 1 Mbp have been found in host-dependent microorganisms, reflecting that the richness of metabolites provided by the host tissue allows for the loss of genes that were previously determined essential (McCutcheon and Moran 2012).

Dufresne et al. (2005) note that bacterial genome size is the result of two counteracting processes: gene duplication (acquisition of new genes) and deletion of non-essential genes. This genomic flux can result in the gene content being altered substantially driving the evolution of bacterial species and the eventual adaptation to new ecological niches (Dufresne et al. 2005). In obligate pathogens, such as *Mycoplasma genitalium*, phytoplasmas and symbionts, gene deletion has prevailed over acquisition resulting in a decrease in size of the

genome (Dufresne et al. 2005). A small population size combined with an inherent bias to delete genes in bacteria ultimately results in genome reduction (Moran 1996; Ochman and Moran 2001; Nilsson et al. 2005; Kuo and Ochman 2009).

There is a clear correlation between genome size and GC content with large genome organisms tending to be GC-rich and reduced genome host obligates tending towards an AT-rich genome (Moran 2002; Bentley and Parkhill 2004). An advantage of an AT-rich genome is that ATP and TTP nucleotides are less energetically expensive (Bentley and Parkhill 2004). Glass et al. (2000) hypothesised that this mutational pressure towards AT enrichment could be a result of the elimination of genes encoding DNA repair enzymes.

One consequence of microbial genome minimisation is that, as rapid sequence evolution occurs, all genes in the genome are affected, ultimately resulting in altered proteins that are predicted to have less stable secondary structures (Fares et al. 2002; van Ham et al. 2003; Fernandez and Lynch 2011). Presymbiotic *Buchnera aphidicola* had a large genome with at least 1800 – 2400 genes (van Ham et al. 2003). After establishment of a symbiotic relationship with the aphid *Baizongia pistacea,* a genome reduction of 65 – 75% occurred, resulting in degenerative rather than adaptive genome evolution (van Ham et al. 2003). The results of degenerate genome evolution include mutational bias, erosion of regulatory systems, and accumulation of mutations affecting protein stability, continued pseudogene formation and further gene loss (van Ham et al. 2003). This thesis will aim to understand whether the effects of genome reduction alter protein stability, function and structure.

## 1.7 Potential Pathogenicity and Virulence Determinants in *Ca.* L. solanacearum

Efforts to understand virulence mechanisms of Liberibacter species has been hindered due to the inability to culture the bacteria in artificial media. Insights into Liberibacter pathogenesis are derived from genome sequences (Duan et al. 2009; Lin et al. 2011; Thompson et al. 2015) and from transcriptomic, proteomic and metabolomics data in response to Liberibacter infection (Fan et al. 2011; Ibanez et al. 2014). An understanding of *C*Lso pathogenicity mechanisms can infer strategic antimicrobial drug design for species specific drug design. Here, the virulence factors and pathogenicity mechanisms of *C*Lso are outlined as this thesis

focuses on the structure and function of select pathogenicity determinants and potential novel antibacterial targets.

### 1.7.1 Putative essential biosynthesis pathways

#### *1.7.1.1 De novo purine biosynthesis*

Purine nucleotides are essential for DNA and RNA production (Le Nours et al. 2011), thus the *de novo* purine biosynthetic pathway is critical for growth and survival (Sassetti et al. 2003). The *C*Lso bifunctional gene, *purH*, encodes a phosphoribosylamino-imidazolecarboxamide formyltransferase cyclohydrolase FAD synthetase (ATIC) which is a component of the *de novo* purine synthesis pathway. The *de novo* purine biosynthesis pathway has been shown to be required for optimal virulence of *S. aureus*, *S. pneumoniae* (Polissi et al. 1998), *Vibrio vulnificus* (Kim et al. 2003), *Salmonella enterica* serovar Typhimurium (McFarland and Stocker 1987), *Yersinia pestis* (Brubaker 1970) and *Bacillus anthracis* (Jenkins et al. 2011). Jenkins et al. (2011) demonstrated that the purine biosynthetic pathway in *Shigella flexneri* is required for full virulence of the bacteria in guinea pigs.

#### *1.7.1.2 Glycolysis*

Recently, Wang et al. (2017b) noted that the genomes of both apiaceous and solanaceous infecting clades of *C*Lso lacked the gene for glucose-6-phosphate isomerase and thus cannot convert α-glucose-6-phosphate to β-fructose-6-phosphate, although a bypass is available via the pentose phosphate pathway. Thus these species rely on β-D-glucose for glycolysis instead of α-D-glucose. β-D-glucose is the substrate for callose and cellulose synthesis in plants: callose deposition is a well-known plant defence mechanism and has been observed *in planta* after infiltration of the *C*Lso flg22 flagellin peptide into plant leaves (Zou et al. 2012). Wang et al. (2017b) speculated that these bacteria use β-D-glucose as a source of carbon for glycolysis in order to reduce the amount of free β-D-glucose that could be used as a substrate for a pattern triggered immunity (PTI) response for plant cell wall fortification (Wang et al. 2017b). The essentiality of the glycolytic pathway for carbon metabolism as well as the use of β-D-glucose for pattern triggered immunity highlight enzymes in the pathway as potential antimicrobial targets.

*1.7.1.3 L-rhamnose synthesis*

L-rhamnose plays an essential role in cell wall synthesis and has been shown to be required for virulence in pathogenic bacteria (Giraud et al. 2000). Vibrio cholera mutants with deletions in the rmlB and and rmlD resulted in the ability of host colonisation, whilst in Streptococcus mutans, inhibition of any gene in the biosynthesis pathway resulted in the inhibition of cell-wall polysaccharide synthesis. Thus, the *L*-rhamnose biosynthetic pathway could play an essential role in *C*Lso pathogenesis and intracellular survival within the host.

### 1.7.2 Lipopolysaccharides

Lipopolysaccharides are a toxic phosphorylated glycolipids present in the cell envelope of gram negative bacteria (Moran 1996) they are crucial for the physical integrity and function of the bacterial outer membrane (Kulshin et al. 1992; Nielsen 1994). Lipopolysaccharides of bacterial pathogens can also trigger plant immune responses such as callose deposition, salicylic acid accumulation and oxidative bursts, whilst symbiotic bacteria are capable of suppressing plant defence responses (Albus et al. 2001; Zipfel and Robatzek 2010). Lipopolysaccharides are known to activate the biosynthetic steps of phenol compounds coumaroyltyramine and feruloyl-tyramine (Newman et al. 2002). *C*Lso results in an increase in phenol levels in infected potato tubers, therefore this could be due to the recognition of lipopolysaccharides by potato cells (Navarre et al. 2009).

### 1.7.3 Salicyclic acid hydroxylase

Pathogen infection of a plant can enhance resistance to subsequent infections through physiological immunity known as systemic acquired resistance (SAR) (Gaffney et al. 1993). This SAR extends to plant tissue that is distant from the initial infection site and can persist for a prolonged period (Gaffney et al. 1993). After pathogen infection but before resistance occurs there is an accumulation of salicylic acid, thus it has been proposed that this salicylic acid serves as a signal molecule for SAR induction (Gaffney et al. 1993; Park et al. 2007).

Li et al. (2017) identified a protein with potential salicylic hydroxylase activity which could convert salicylic acid into catechol which is present in all Liberibacter species and encoded *sahA*. Thus, the presence of this protein could allow the organism to evade plant defences by

modulating overall defence of its host by degrading salicylic acid and suppressing plant defences (Dalio et al. 2017; Wang 2017b).

### 1.7.4 *Ca*. L. solanacearum encodes a type I secretion system

Bacterial secretion systems are commonly used by many bacterial pathogens for the delivery of virulence factors to the extracellular space, or alternatively directly into host cells (Baron and Coombes 2007). However, the *Ca*. Liberibacter species lack the classic mechanisms for protein secretion. Lin et al. (2011) determined there was no evidence of a type II or III bacterial secretion system in *C*Lso, however, the organism does possess all the elements of a type I secretion system. Most of the Sec-SRP system was also present, although the *SecB* chaperone gene was not present. Some components of the type III and type IV systems were present, the genome did not contain a type II system (Lin et al. 2009). Yao et al. (2016) reported the presence of 16 putative type IV effector genes, four of which (pilus assembly proteins) were differentially expressed between the *C*Lso A and B haplotypes, although the presence of a type IV system was not noted (Yao et al. 2016). The type IV pilus secretion system enables the transport of virulent proteins, bacterial conjugation and DNA uptake or release which in turn promotes the exchange of genetic material additionally, in host recognition and attachment, invasion and biofilm formation. (Rego et al. 2010; Yao et al. 2016).

The *C*Lso genome maintains genes encoding a complete type I secretion system with genes such as TolC, HlyD and PrtD, as well as the outer membrane porin protein (OmpA) and the flagellin regulatory protein (FlbT) which are required for the secretion of toxins and efflux of plant defence compounds (Lin et al. 2011; Lin et al. 2013). Furthermore, putative type I secretion system substrates such as serralysin and haemolysin have also been identified in the Liberibacter genomes (Duan et al. 2009: Wang 2017b).

Repeats-in-ToXin (RTX) toxins are important virulence factors in Gram-negative bacteria (Frey 1995; Lally et al. 1999). The *Ca*. L solanacearum genomes has genes encoding orthologs of *hlyD* and *prtD* which are clustered together with a putative RTX toxin (Lin et al. 2011) and are potentially secreted by the type I secretion system.

The serralysin metalloproteases are members of the RTX family of toxins and are associated with virulence in multiple nodes of infection (Zhang et al. 2015). *Serratia marcescens*, a human pathogen, is associated with pneumonia, urinary tract infection, corneal keratitis, meningitis and endocarditis (Zhang et al. 2015). Although the pathogen produces multiple virulence factors to promote adherence and facilitate infection, the serralysin metalloprotease is the most abundant extracellular virulence factor produced. Ravindran et al. (2017) determined the presence of a serralysin like protein in *Ca.* Liberibacter solanacearum. Ravindran et al. (2017) speculated that the proximity of the *C*Lso serralysin-like gene to divergently transcribed genes encoding the type I secretion system could be a significant virulence factor. Gene expression analysis of the serralysin-like gene showed a continued increase in expression as infection with *C*Lso progressed (Ravindran et al. 2017).

Interestingly, the *Ca*. Liberibacter species have retained the elements of the complete secretory (extracellular secretion) machinery (SEC). The SEC translocase provides a major pathway of protein translocation from the cytosol across the cytoplasmic membrane to the environment in bacteria and is essential for bacterial viability (Mori and Ito 2001; Wang 2017b). SEC was found to have a prominent role in the secretion of major bacterial virulence determinants in *Staphylococcus aureus* with deletion of the bifunctional SECDF protein resulting in reduced resistance, alteration of cell separation and change in expression of virulence factors (Quiblier et al. 2013). Furthermore, the SEC machinery secretes vital virulence factors in plant-pathogenic bacteria (Wang 2017b). In phytoplasmas, the virulence factors SAP54, SAP11, and Tengu are secreted into plant cells (Hoshi et al. 2009; MacLean et al. 2011; Sugio et al. 2011; Wang 2017b). Signal peptide containing extracytoplasmic proteins (SPEP) have been are potential substrates for the SEC machinery with 171 SPEP predicted in *C*Lso (Prasad et al. 2016; Wang 2017b). The SEC protein translocase belongs to the resistance nodulation division (RDN) family of multi drug exporters, which are known to mediate resistance to various substances and prevent the accumulation of antibiotics within the cell (Quiblier et al. 2013).

### 1.7.5 Flagellin acts as a pathogen associated molecular pattern

Haapalainen (2014) concluded that *Ca.* Liberibacter species are lacking the common pathogenicity and virulence determinants which could be recognised as a pathogen associated molecular pattern (PAMP) by plants, and thus avoid eliciting a strong defence response *in planta* during infection. Fagen et al. (2014b) compared the genomes of uncultured and cultured *Liberibacter* species and noted that there were no significant differences in terms of pathogenicity or virulence determinant, concluding that *L. crescens* may be a plant pathogen on an unidentified plant host. Flagella are essential structures for pathogenic bacteria, they not only confer motility but also promote adherence by acting as bacterial adhesins, promote bacterial biofilm formation and trigger host responses (Duan et al. 2013). The *Ca*. L. asiaticus genome was found to retain the entire flagellum encoding gene cluster including a flagellin and hook associated protein (Fla protein) that contained a 22 amino acid conserved domain (*flg22*) (Zou et al. 2012). Like *Ca*. L. asiaticus, the *C*Lso genome contains the same number of open reading frames that are essential for the structure and assembly of a flagellum (Duan et al. 2009). Yao et al. 2016 found that the *C*Lso flagellin gene was highly expressed in *B. cockerelli.* Zou et al. (2012) demonstrated that a functional *Ca*. L. asiaticus flagellin acted as a pathogen associated molecular pattern (PAMP) by inducing callose deposition, but did not result in plant cell death. However, microscopic analysis of *Ca.* L. asiaticus infected citrus showed accumulation of callose in sieve plates (Koh et al. 2012). Excessive callose deposition in the phloem plasmodesmata could interrupt photoassimilate distribution and cause starch over accumulation in leaf chloroplasts (Koh et al. 2012).

Recently, Wang et al. (2017b) noted that the genomes of both apiaceous and solanaceous infecting clades of *C*Lso lacked the gene for glucose-6-phosphate isomerase and thus cannot convert α-glucose-6-phosphate to β-fructose-6-phosphate, although a bypass is available via the pentose phosphate pathway. Thus these species rely on β-D-glucose for glycolysis instead of α-D-glucose. β-D-glucose is the substrate for callose and cellulose synthesis in plants: callose deposition is a well-known plant defence mechanism and has been observed *in planta* after infiltration of the *C*Lso flg22 flagellin peptide into plant leaves (Zou et al. 2012). Wang et al. (2017b) speculated that these bacteria use β-D-glucose as a source of carbon for glycolysis

in order to reduce the amount of free β-D-glucose that could be used as a substrate for a pattern triggered immunity (PTI) response for plant cell wall fortification (Wang et al. 2017b).

### 1.7.6 Prophage regions could be responsible for virulence variation between haplotypes

Annotation and comparison of the *C*Lso genomes identified the presence of prophage domains. The presence of an additional prophage domain in NZ1 increased the genome size of the organism to 1.31 Mbp in comparison to the 1.26 Mbp size of the ZC1 genome (Thompson et al. 2015). The presence and variation of these prophage regions within the *C*Lso genomes may be important in understanding the variation of virulence that is observed between haplotypes since it has been observed that the genotype of *Ca.* Liberibacter asiaticus appear to correlate with disease development in plant hosts (Zhou et al. 2013). Lin and Gudmestad (2013) and Thompson et al. (2015) speculated that differences in prophage regions between *C*Lso haplotypes may result in changes in virulence and that these changes may have contributed to the variation in severity of Zebra Chip disease observed in New Zealand compared to the United States.

Prophages are bacteriophage genomes integrated into the host genome (Winstanley et al. 2009). Genomic comparison of the *C*Lso ZC1 (haplotype B) and NZ1 (haplotype A) showed major differences in the genomic locations and composition of the prophage regions with ZC1 having two prophage regions and three prophage regions present in NZ1 (Thompson et al. 2015). Prophages present in *Ca*. L. asiaticus are thought to contribute to suppressing plant defence systems (Jain et al. 2015) due to the prophage encoding a reactive oxygen species scavenging peroxidase. This peroxidase downregulates RbohB preventing the hydrogen peroxide mediated defence signal in plants and as a result *Ca*. L. asiaticus evades early detection in the plant (Jain et al. 2015). The diversity and number or prophage regions could be responsible for the variation in virulence between haplotypes and could result in the variation in disease severity of Zebra Chip observed (Thompson et al. 2015).

## 1.8 Limitations in *Ca.* Liberibacter research

Insights into Liberibacter pathogenesis and proteins have mainly been derived from genomic analysis (Duan et al. 2009; Lin et al. 2011; Leonard et al. 2012; Fagen et al. 2014; Wulff et al. 2014; Chika et al. 2017; Levy et al. 2017; Huot et al. 2018) and host transcriptomic and metabolomic responses to Liberibacter infection (Fan et al. 2011; Slisz et al. 2012; Mafra et al. 2013; Ibanez et al. 2014). There is little knowledge of the structure and function of *C*Lso proteins which has hindered the progress in the development of an antimicrobial. Currently, studies on recombinant expression of *Ca*. Liberibacter proteins highlight the difficulties of purifying active *C*Lso proteins. Ravindran et al. (2018) purified the serralysin-like protein (haplotype LsoA KJZ82330.1 and haplotype LsoB WP_013461860 LsoA) from *Candidatus* Liberibacter solanacearum haplotypes LsoA and LsoB (HenneA and ZC1) under denaturing conditions. However, the purified enzyme had no measurable protease activity, leading to the conclusion that either the putative gene does not encode a functional enzyme, or that the protein incorrectly folds during the purification and refolding process (Ravindran et al. 2018). Liu et al. (2017) purified six proteins from *Ca*. L. asiaticus psy62 under native and denaturing conditions; however, the activity of these proteins was not studied. Furthermore, Lau et al. (2011) purified the partial outer membrane protein (Omp) of *Candidatus* L. asiaticus under denaturing conditions, however, the protein was not characterized. The inability to culture *C*Lso, along with no current characterisation of functional proteins has hindered the understanding of *C*Lso biology and pathogenicity.

## 1.9 Summary

Zebra Chip disease has not only caused substantial economic losses in commercial potato production, and threatens international food security. Potatoes are the fourth most consumed food crop producing more food per unit of water than any other food crop (CIP 2010). The psyllid vector alone caused production losses to producers of NZD$43 million in 2008 and 2009 in New Zealand (Plant Health Australia 2010). Total losses and psyllid control costs are estimated at NZD$200 million since 2008 (CRC Plant Biosecurity 2016). Due to the damaging effects of this disease, it is imperative to gain further insight into the biology of the causal organism, *C*Lso for future development of antibacterials against the bacterium.

How species in the *Liberibacter* genus infect plants and cause disease is unclear. Potential pathogenicity and virulence determinants identified to date have been analysed via comparative genomics of bacterial genomes assembled from metagenomic sources (Duan et al. 2009; Lin et al. 2011; Hartung et al. 2011; Leonard et al. 2012; Fagen et al. 2014a; Fagen et al. 2014b). Although these methods have provided understanding of the genes present within the organism, understanding the relative importance of each of the possible pathogenicity genes and their structure and function relationship in the different plant host species will be challenging and critical to the development of a successful new control technology for this emerging plant pathogen genus.

## 1.10 Aims and objectives of the thesis

The overall aim of this research is to identify, produce and investigate proteins that could contribute to pathogenicity of *C*Lso and proteins that could be potential antibacterial targets. Given the unculturable nature of the pathogen, insights into the biology of the organism and structural and enzymatic properties of encoded proteins are limited. Probing the structure of essential proteins in this organism could inform future antibacterial drug design and expand our current knowledge of how this bottlenecked, reduced genome organism survives and confers disease to host plants.

*Chapter 2* investigates the gene expression of a number of potential pathogenicity genes and their differential expression in the psyllid vector (*Bactericera cockerelli*) compared to *in planta*. Gene expression data is presented and provided insight into which genes could possibly play a role in pathogenicity and for subsequent protein characterisation.

*Chapter 3* investigates structural insights into the bifunctional adenyltransferase/IMP cyclohydrolase, a putative pathogenicity factor of *C*Lso. Biophysical characterization is presented here.

*Chapter 4* addresses the difficulties in recombinant expression and purification of proteins from *Candidatus* Liberibacter solanacearum. An approach to express soluble, functional *C*Lso proteins was established by mimicking *in vivo* conditions of high chaperone expression in *E.*

*coli* proteins *in vitro*. This demonstrates that this approach produces functional dihydrodipicolinate synthase and pyruvate kinase both of which exhibit the canonical tetrameric oligomeric state seen in other organisms.

*Chapter 5* presents the structure of *C*Lso pyruvate kinase, a key protein in glycolysis. The low resolution structure presented here, investigates the changes in active and allosteric site of the protein compared to other bacterial homologues in order to determine whether changes in these sites result in attenuated activity for *C*Lso pyruvate kinase

*Chapter 6* investigates the structure and catalytic mechanism of *C*Lso dihydrodipicolinate synthase. Dihydrodipicolinate synthase (DHDPS) is the rate limiting step in *L*-lysine biosynthesis and a potential drug target. The structure of *C*Lso dihydrodipicolinate synthase is the first acquired for any protein of *C*Lso. I solved the structure in three conformations: with lysine bound in the allosteric site, with pyruvate, and with SSA (an analogue of ASA) and pyruvate.

## 1.11 References

Abad, J.A., Bandla, M., French-Monar, R.D., Liefting, L.W., and Clover, G.R.G. 2009. First report of the detection of '*Candidatus* Liberibacter' species in Zebra Chip disease-infected potato plants in the United States. Plant Disease 93:108.

Albus, U., Baier, R., Puhler, A., and Niehaus, K. 2001. Suppression of an elicitor induced oxidative burst reaction in *Medicago sativa* cell cultures by *Sinorhizobium meliloti* lipopolysaccharides. New Phytology 151:597-606.

Alfaro-Fernandez, A., Siverio, F., Cebrian, M., Villaescusa, F., and Font, M. 2011. *Candidatus* Liberibacter solanacearum associated with *Bactericera trigonica* affected carrrots in the Canary Islands. Plant Disease 96:581.

Baron, C., and Coombes, B. 2007. Targeting bacterial secretion systems: benefits of disarmament in the microcosm. Infectious Disorders Drug Targets 7:19–27.

Bastolla, U., and Demetrius, L. 2005. Stability constraints and protein evolution: the role of chain length, composition and disulphide bonds. Protein Engineering Design and Selection 18: 405-415.

Bentley, S.D., and Parkhill, J. 2004. Comparative genomic structure of prokaryotes. Annual Review of Genetics 38:771–792.

Brenner, S.E. 1999. Errors in genome annotation. Trends in Genetics. 15(4):132-3.

Brown, J.S., Gilliand, S.M., and Holden, D.W. 2001. A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporters involved in iron uptake and virulence. Molecular Microbiology 40:572-585.

Brubaker, R.R. 1970. Interconversion of purine mononucleotides in *Pasteurella pestis*. Infection and Immunity 1:446-54.

Buchman, J.L., Fisher, T.W., Sengoda, V.G., and Munyaneza, J.E. 2012. Zebra Chip progression: from inoculation of potato plants with Liberibacter to development of disease symptoms in tubers. American Journal of Potato Research 89:159–168.

Buchman, J.L., Sengoda, V.G., and Munyaneza, J.E. 2011. Vector transmission efficiency of liberibacter by *Bactericera cockerelli* (Hemiptera: Triozidae) in Zebra Chip potato disease: effects of psyllid life stage and inoculation access period. Journal of Economic Entomology 1045:1486–1495.

Butler, C.D., and Trumble, J.T. 2012. The potato psyllid, (Sulc) (Hemiptera: Triozidae): life history, relationship to plant diseases, and management strategies. Terrestrial Arthropod Reviews 5:87–111.

Casteel, C.L., Hansen, A.K., Walling, L.L., and Paine, T.D. 2012. Manipulation of plant defense responses by the tomato psyllid (*Bactericerca cockerelli*) and its associated endosymbiont *Candidatus* Liberibacter psyllaurous. PLoS ONE 7:4.

Chen, C., Chen, X., Xie, T., Hatting, J.L., Yu, X., Ye, S., Wang, Z., and Shentu, X. 2016. Diverse bacterial symbionts of insect-pathogenic fungi and possible impact on the maintenance of virulence during infection. Symbiosis 69:47-58.

Coletta-Filho, H.D., Targon, M.L.P.N., Takita, M.A., De Negri, J.D., Pompeu, J., Machado, M.A., Amaral, A.M., and Muller, G.W. 2004. First report of the causal agent of huanglongbing *Candidatus* Liberibacter asiaticus in Brazil. Plant Disease 88:1382.

CRC Plant Biosecurity. 2016. Case study: Zebra Chip. PBCRC Publishing. http://www.pbcrc.com.au/news/2016/pbcrc/case-study-zebra-chip. Accessed 2nd January 2017.

Crosslin, J.M., Munyaneza, J.E., Brown, J.K., and Liefting, L.W. 2010. Potato Zebra Chip disease; a phytopathological tale. Plant Health Progress 11(1): 33.

Davidson, A.L., Dassa, E., Orelle, C., and Chen, J. 2008. Structure, function and evolution of bacterial ATP-binding cassette systems. Microbiology and Molecular Biology Reviews 72:317-364.

Dalio, R.J.D., Magalhaes, D.M., Rodrigues, C.M., Arena, G.D., Oliveira, T.S., Souza-Neto, R.R., Picchi, S.C., Martins, P.M.M., Santos, P.J.C., Maximo, H.J., Pacheco, I.S., De Souza, A.A., and Machado, M.A. 2017. PAMPS, PRR's effectors and R-genes associated with citrus-pathogen interactions. Annals of Botany 119:749-774.

Dehbashi, S., Pourmand, M.R., Mashhadi, R. 2016. Characterization of Afb, a novel bifunctional protein in *Streptococcus agalactiae*. Iranian Journal of Microbiology 8:73-79.

Devos, D., and Valencia, A. 2001. Intrinsic errors in genome annotation. TRENDS in Genetics 17(8):429-31.

Duan, Y., Gottwald, T., Zhou, L., Gabriel, D. 2008. First report of dodder transmission of '*Candidatus* Liberibacter asiaticus' to tomato (*Lycopersicon esculentum*). Plant Disease 92:831.

Duan, Y., Zhou, L., Hall, D.G., Li, W., Doddapaneni, H., Lin, H., Liu, L., Vahling, C.M., Gabriel, .W., Williams, K.P., Dickerman, A., Sun, Y., Gottwald, T. 2009. Complete genome sequence of citrus huanglongbing bacterium, '*Candidatus* Liberibacter asiaticus' obtained through metagenomics. Molecular Plant-Microbe Interactions 228:1011–1020.

Duan, Q., Zhou, M., Zhu, L., and Zhu, G. 2013. Flagella and bacterial pathogenicity. Journal of Basic Microbiology 53:1-8.

Dufresne, A., Garczarek, L., and Partensky, F. 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biology 6:14.

Fagen, J.R., Leonard, M.T., Coyle, J.F., Mccullough, C.M., Davis-Richardson, A.G., Davis, M.J., and Triplett, E.W. 2014a. *Liberibacter crescens* the first cultured member of the genus liberibacter. International Journal of Systematic and Evolutionary Microbiology 64:2461–2466.

Fagen, J.R., Leonard, M.T., McCullough, C.M., Edirisinghe, J.N., Henry, C.S., Davis, M.J., and Triplett, E.W. 2014b. Comparative genomics of cultured and uncultured strains suggests genes essential for free-living growth of Liberibacter. PLoS ONE 9:1–11.

Fagen, J.R. 2014c. Comparative study of Liberibacter species divergent in culturability and virulence. Dissertation, University of Florida.

Fares, M.A., Ruiz-González, M., Moya, A., Santiago, E.F., and Barrio, E. 2002. Endosymbiotic bacteria: *groEL* buffers against deleterious mutations. Nature 417:398.

Feng, Y., Chin, C.Y., Chakravartty, V., Gao, R., Crispell, E.K., Weiss, D.S., Cronan, J.E. 2015. The atypical occurrence of two biotin protein ligases in *Francisella novicida* is due to distinct roles in virulence and biotin metabolism. mBio 60:591-615.

Fernández, A., and Lynch, M. 2011. Non-adaptive origins of interactome complexity. Nature 474:502–505.

Fisher, T.W., Vyas, M., He, R., Nelson, W., Cicero, J.M., Soderlund, C.A., Gang, D.R., and Brown, J.K. 2014. Comparison of potato and asian citrus psyllid adult and nymph transcriptomes identified vector transcripts with potential involvement in circulative, propagative Liberibacter transmission. Pathogens 3:875–907.

Fitzgerald, B. 2017. West Australian potato disease threats stunts trade as growers warned spread almost inevitable. Australian Broadcasting Company. http://www.abc.net.au/news/rural/2017-03-31/wa-seed-industry-frustrated-over-biosecurity-trade-halt/8401726. Accessed 30 July 2017.

Frey, J. 1995. Virulence in *Actinobacillus pleuropneumoniae* and RTX toxins. Trends in Microbiology 3:257-261.

Gaffney, T., Friedrich, L., Vernooij, B., Negrotto, D., Nye, G., Uknes, S., Ward, E., Kessmann, H., and Ryals, J. 1993. Requirement of salicylic acid for the induction of systemic acquired resistance. Science 261:754-756.

Garnier, M., Danel, N., and Bové, J.M. 1984. The greening organism is a gram negative bacterium. Proceedings of 9th Conference IOCV 115-124.

Garnier, M., Jagoueix-Eveillard, S., Cronjem, P.R., Le Roux, H.F., and Bove, J.M. 2000. Genomic characterization of a liberibacter present in an ornamental rutaceous Tree,

*Calodendrum capense* , in the western cape province of South Africa proposal of *Candidatus* Liberibacter africanus subsp. capensis'. International Journal of Systematic and Evolutionary Microbiology 50:2119–2125.

Gill, G. 2006. Tomato psyllid detected in New Zealand. Biosecurity 69:10–11.

Glaser, P., Sakamoto, H., Bellalou, J., Ullmann, A., and Danchin, A. 1988. Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*. EMBO Journal 7:3997-4004.

Glass, J.I., Lefkowitz, E.J., Glass, J.S., Heiner, C.R., Chen, E.Y., and Cassell, G.H. 2000. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. Nature 407:757–762.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Research 38:695-699.

Haapalainen, M. 2014. Biology and epidemics of *Candidatus* Liberibacter species, psyllid-transmitted plant-pathogenic bacteria. Annals Applied Biology 165:172–198.

Hansen, A.K., Trumble, J.T., Stouthamer, R., and Paine, T.D. 2008. A new huanglongbing species, '*Candidatus* Liberibacter psyllaurous' found to infect tomato and potato, is vectored by the psyllid *Bactericera cockerelli* (Sulc). Applied and Environmental Microbiology 74:5862–5865.

Hartung, J.S., Shao, J., and Kuykendall, D.L. 2011. Comparison of the *Candidatus* Liberibacter asiaticus genome adapted for an intracellular lifestyle with other members of the Rhizobiales. PLOS One 6:23289.

Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., Komatsu, K., Kagiwada, S., Yamaji, Y., and Namba, S. 2009. A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. Proc Natl Acad Sci USA 106:6416-6421.

Jagoueix, S., Bove, J.M., and Garnier, M. 1994. The phloem-limited bacterium of greening disease of citrus is a member of the alpha subdivision of the proteobacteria. International Journal of Systematic Bacteriology 44:379–386.

Jain, M., Fleites, L.A., and Gabriel, D.W. 2015. Prophage encoded peroxidase in *Candidatus* Liberibacter asiaticus is a secreted effector that suppresses plant defences. Molecular Microbe Interactions 28:1330-1337.

Jenkins, A., Cote, C., Twenhafel, N., Merkel, T., Bozue, J., and Welkos, S. 2011. Role of purine biosynthesis in *Bacillus anthracis* pathogenesis and virulence. Infection and Immunity 79:153-166.

Jensen, J.H. 1939. Psyllid yellows in Nebraska 1938. Plant Disease Reports 23:35-26.

Johnston, J.M., Arcus, V.L., Morton, C.J., Parker, M.W., and Baker, E.N. 2003. Crystal structure of a putative methyltransferase from *Mycobacterium tuberculosis*: misannotation of a genome clarified by protein structural analysis. Journal of bacteriology 185(14):4057-65.

Jorgensen, N., Butler, R.C., and Vereijssen, J. 2013. Biorational insecticides for control of the tomato potato psyllid. New Zealand Plant Protection 66:333–340.

Karp, P.D. 1998. What we do not know about sequence analysis and sequence databases. Bioinformatics 14(9):753-4.

Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, E.K., Shin, D.H., Chung, S.S., Choy, H.E., Progulske-Fox, A., Hillman, J.D., Handfield, M., and Rhee, J.H. 2003. Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. Infection and Immunity 71:5461-5471.

Koh, E.J., Zhou, L., Williams, D.S., Park, J., Ding, N., Duan, Y., and Kang, B.H. 2012. Callose deposition in the phloem plasmodesmata and inhibition of phloem transport in citrus leaves infected with *Candidatus* Liberibacter asiaticus. Protoplasma 249:687-697.

Kulshin, V.A., Zahringer, U., Lindner, B., Frasch, C.E., Tsai, C.M., Dmitriev, B.A., Rietschel, E.T. 1992. Structural characterization of the lipid A component of pathogenic *Neisseria meningitidis*. Journal of Bacteriology 174:1793-1800.

Kuo, C.H., and Ochman, H. 2009. Deletional bias across the three domains of life. Genome Biological Evolution 1:145–152.

Lacey, L.A., Liu, T.X., Buchman, J.L., Munyaneza, J.E., Goolsby, J.A., and Horton, J.R. 2011. Entomopathogenic fungi (hypocreales) for control of potato psyllid, *Bactericera cockerelli* (Šulc) (Hemiptera: Triozidae) in an area endemic for Zebra Chip disease of potato. Biological Control 56:271–278.

Lafleche, D., and Bove, J.M. 1970. Mycoplasmes dans les agrumes atteints de ‘greening’, de ‘stubborn’ ou de maladies similaires. Fruits 25:455–465.

Lai, K.K., Davis-Richardson, A.G., Dias, R., and Triplett, E.W. 2016. Identification of the genes required for the culture of *Liberibacter crescens*, the closest cultured relative of the uncultured liberibacter plant pathogens. Frontiers in Microbiology 7:041–11.

Lally, E.T. 1999. The interaction between RTX toxins and target cells. Trends Microbiology 7:356-361.

Lapage, S.P., Sneath, P.H.A., Lessel, E.F., Skerman, V.B.D., Seeliger, H.P.R., and Clark, W.A. 1992. International code of nomenclature of bacteria: bacteriological code 1990 revision. ASM Press, Washington DC.

Leonard, M.T., Fagen, J.R., Davis-Richardson, A.G., and Davis, M.J. 2012. Complete genome sequence of *Liberibacter crescens* BT-1. Standards in Genomic Sciences 7:271–283.

Lercher, B.Y., and Pal, C. 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol Bio Evol 25:559-567.

Li, A., Geng, J., Cui, D., Shu, C., Zhang, S., Yang, J., Xing, J., Wang, J., Ma, F., and Hu, S. 2011 Genome sequence of *Agrobacterium tumefaciens* strain F2, a bioflocculant producing bacterium. J Bacteriol 193:5531.

Li, J., Pang, Z., Trivedi, P., Zhou, X., and Ying, X. 2017. *Candidatus* Liberibacter asiaticus encodes a functional salicylic acid (SA) hydroxylase that degrades SA to suppress plant defences. Molecular Plant Microbe Interactions 30:620-630.

Li, W., Cong, Q., Pei, J., Kinch, L.N., and Grishin, N.V. 2012. The ABC transporters in *Candidatus* Liberibacter asiaticus. Proteins 80:2614-2628.

Li, W., Abad, J.A., French-Monar, R.D., Rascoe, J., Wen, A., Gudmestad, N.C., Secor, G.A., Lee, I.M., Duan, Y., and Levy, L. 2009. Multiplex real-time PCR for detection, identification and quantification of '*Candidatus* Liberibacter solanacearum' in potato plants with Zebra Chip. Journal of Microbiological Methods 78:59–65.

Liefting, L.W., Ward, L.I., Shiller, J.B., and Clover, G.R.G. 2008a. A new *Candidatus* Liberibacter species in *Solanum betaceum* (tamarillo) and *Physalis peruviana* (cape gooseberry) in New Zealand. Plant Disease 92:1588.

Liefting, L.W., Perez-Egusquiza, C., and Clover, G.R.G. 2008b. A new '*Candidatus* Liberibacter' species in *Solanum tuberosum* in New Zealand. Plant Disease 92:1474.

Liefting, L.W., Weir, B.S., Pennycook, S.R., and Clover, G.R.G. 2009. *Candidatus* Liberibacter solanacearum associated with plants in the family Solanaceae. International Journal of Systematic and Evolutionary Microbiology 59:2274-2276.

Lin, H., Lou, B., Glynn, J.M., Doddapaneni, H., Civerolo, E.L., Chen, C., Duan, Y., Zhou, L., Vahling, C.M. 2011. The complete genome sequence of '*Candidatus* Liberibacter Solanacearum ', the bacterium associated with potato Zebra Chip disease. PLOS One 6:19135.

Lin, H., and Gudmestad, N.C. 2013. Aspects of pathogens genomics, diversity, epidemiology, vector dynamics and disease management for a newly emerged disease of potato: Zebra Chip. Phytopathology 103:524-37.

MacLean, A.M., Sugio, A., Makarova, O.V., Findlay, K.C., Grieve, V.M., and Toth, R. 2011. Phytoplasma effectors SAP54 induces indeterminate leaf-like flower development in Arabidopsis plants. Plant Physiology 157:831-841.

Mathieu, M., Debousker, G., Vincent, S., Viviani, F., Bamas-Jacques, N., and Mikol, V. 2005. *Escherichia coli* FolC structure reveals an unexpected dihydrofolate binding site providing an attractive target for anti-microbial therapy. Journal of Biological Chemistry 280:18916–18922.

McCutcheon, J.P., and Moran, N.A. 2012. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10:13-26.

McFarland, W.C., and Stocker, B.A.D. 1987. Effect of different purine auxotrophic mutations on mouse-virulence of a vi-positive strain of *Salmonella* Dublin and of two strains of *Salmonella Typhimurium*. Microbial Pathogenesis 3:129-141. doi: 10.1016/0882-4010(87)90071-4.

McWilliam, H., Li, W., Uludag, M., Squizzati, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. 2013. Analysis tool web services from the EMBL-EBI. Nucleic Acids Research 41:597-600.

Mendonca, A.G., Alves, R.J., and Pereira-Leal, J.B. 2011. Loss of genetic redundancy in reductive genome evolution. PLoS Computational Biology 7:e1001082.

Miles, G.P., Samuel, M.A., Chen, J., Civerolo, E.L., and Munyaneza, J.E. 2010. Evidence that cell death is associated with Zebra Chip disease in potato tubers. American Journal of Potato Research 87:337–349.

Moran, N.A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci USA 93:2873–2878.

Moran, N.A. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108:583–586.

Morris, J., Shiller, J., Mann, R., Smith, G., Yen, A., and Rodoni, B. 2017. Novel *Candidatus* Liberibacter species identified in the Australian eggplant psyllid *Acizzia solanicola*. Micorbial Biotechnology 10:833-844.

Mori, H., and Ito, K. 2001. The Sec protein-translocation pathway. Trends in Microbiology 9:494-500.

Munyaneza, J.E. 2012. Zebra Chip disease of potato : biology, epidemiology and management. American Journal of Potato Journal 89:329–350.

Munyaneza, J.E., Buchman, J.L., Sengoda, V.G., Fisher, T.W., and Pearson, C.C. 2011. Susceptibility of selected potato varieties to Zebra Chip disease. American Journal of Potato Research 88:435.

Munyaneza, J.E., Buchman, J.L., Upton, J.E., Goolsby, J.A., Crosslin, J.M., Bester, G., Miles, G.P., and Sengoda, V.G. 2008. Impact of different potato psyllid populations on Zebra Chip disease incidence, severity, and potato yield. Subtropical Plant Science 60:27–37.

Munyaneza, J.E., Fisher, T.W., Sengoda, V.G., Garczynski, S.F., Nissinen, A., and Lemmetty, A. 2010a. First report of '*Candidatus* Liberibacter solanacearum' associated with psyllid-affected carrots in Europe. Plant Disease 94:639-639.

Munyaneza, J.E., Fisher, T.W., Sengoda, V.G., Garczynski, S.F., Nissinen, A., Lemmetty, A. 2010b. Association of '*Candidatus* Liberibacter Solanacearum' with the psyllid, *Trioza apicalis* (Hemiptera: Triozidae) in Europe. Journal of Economic Entomology 103:1060–1070.

Munyaneza, J.E., Goolsby, J.A., Crosslin, J.M., and Upton, J.E. 2007. Further evidence that Zebra Chip potato disease in the Lower Rio Grande Valley of Texas is associated with *Bactericera cockerelli*. Subtropical Plant Science 59:30-37.

Navarre, D.A., Shakya, R., Holden, J., and Crosslin, J.M. 2009. LC-MS analysis of phenolic compounds in tubers showing Zebra Chip symptoms. American Journal of Potato Research 86:88-95.

Nelson, W.R., Fisher, T.W., and Munyaneza, J.E. 2011. Haplotypes of '*Candidatus* Liberibacter solanacearum' suggest long-standing separation. European Journal of Plant Pathology 130:5–12.

Newman, M.A., von Roepenack-Lahaye, E., Parr, A., Daniels, M.J., and Dow, J.M. 2002. Prior exposure to lipopolysaccharides potentiates expression of plant defences in response to bacteria. The Plant Journal 29:487-495.

Nielsen, H., Birkholz, S., Andersen, L.P., and Moran, A.P. 1994. Neutrophil activation by Helicobacter pylori lipopolysaccharides. Journal of Infectious Diseases 170:135-139.

Nilsson, A.I., Koskiniemi, S., Eriksson, S., Kugelberg, E., Hinton, J.C.D., and Andersson, D.I. 2005. Bacterial genome size reduction by experimental evolution. Proceedings of the National Academy of Sciences of the United States of America 102:12112-12116.

O'Connell, D.M., Wratten, S.D., Pugh, A.R., and Barnes, A.M. 2012. New species association biological control, two coccinellid species and an invasive psyllid pest in New Zealand. Biological Control 62:86–92.

Ochman, H., and Moran, N.A. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292:1096–1099.

Ochman, H., and Kelkar, Y.D. 2013. Genome reduction promotes increase in protein functional complexity in bacteria. Genetics 193:303-307.

Ogden, S.C. 2011. Tomato potato psyllid and liberibacter in New Zealand-impact and research programme overview. Proceedings of the 11th Annual Zebra Chip Reporting Session 6–9.

Park, S.W., Kaimoyo, E., Kumar, D., Mosher, S., and Klessig, D.F. 2007. Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. Science 318:113-116.

Pereira, S., Zille, A., Micheletti, E., Moradas-Ferreira, P., De Phillipis, R., and Tamagnini, P. 2009. Complexity of cyanobacterial exopolysaccharideas: composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. FEMS Microbiology Reviews 33:917-941.

Plant Health Australia. 2010. Threat Specific Contingency Plan: Zebra Chip Complex. Industry Biosecurity Plan for the Potato Industry. http://www.planthealthaustralia.com.au/wp-content/uploads/2013/03/Zebra-chip-CP-2011.pdf. Accessed 14th February 2017.

Pletsch, D.J. 1947. The potato psyllid *Paratrioza cockerelli* (Sulc) in biology and control. EPPO Bulletin 446:95.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. 1998. Large scale identification of virulence genes from *Streptococcus pneumoniae*. Infect Immun 66:5620-9.

Prager, S.M., Wallis, C., and Trumble, J.T. 2015. Indirect effects of one plant pathogen on the transmission of a second pathogen and the behavior of its potato psyllid vector. Environmental Entomology 44:1065–1075.

Prasad, S., Xu, J., Zhang, Y., and Wang, N. 2016. SEC-translocon dependent extracytoplasmic proteins of Candidatus Liberibacter asiaticus. Frontiers in Microbiology 7:1-9.

Purushothaman, S., Gupta, G., Srivastava, R., Ramu, V.G., and Surolia, A. 2008Ligand specificity of group I biotin protein ligase of *Mycobacterium tuberculosis*. PLoS ONE 3:1–12.

Quiblier, C., Seidl, K., Roschitzki, B., Zinkernagel, A.S., Berger-Bachi, B., and Senn, M.M. 2013. Secretome analysis defines the major role of SECDF in *Staphylococcus aureus*. PLos One 8:e63513.

Raddadi, N., Gonella, E., Camerota, C., Pizzinat, A., Tedeschi, R., Crotti, E., Mandrioli, M., Bianco, P.A., Daffonchio, D., and Alma, A. 2011. *Candidatus* Liberibacter europaeus that is associated with and transmitted by the psyllid *Cacopsylla pyri* apparently behaves as an endophyte. Environmental Microbiology 13:414–426.

Ravindran, A., Saenkham, P., Gad-Levy, J., Tamborindeguy, C., Lin, H., Gross, D., Pierson, E.A. 2018. Characterization of the serralysin-like gene of *Candidatus* Liberibacter solanacearum associated with potato Zebra Chip disease. Phytopathology 108(3): 327–335.

Rego, A.T., Chandran, V., and Waksman, G. 2010. Two step and one step secretion mechanisms in gram negative bacteria: contrasting the type IV secretion system and the chaperone usher pathway of pilus biogenesis. Biochemical Journal 425:475-488.

Richards, H.L. 1928. A new and destructive disease of the potato in Utah and its relation to the potato psylla. Phytopathology 18:140–141.

Sandanayaka, W.R.M., Moreno, A., Tooman, L.K., Page-Weir, N.E.M., and Fereres, A. 2014. Stylet penetration activities linked to the acquisition and inoculation of *Candidatus* Liberibacter solanacearum by its vector tomato potato psyllid. Entomologia Experimentalis et Applicata. 151:170–181.

Schnoes, A.M., Brown, S.D., Dodevski, I., and Babbitt, P.C. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS computational biology 5(12):e1000605.

Secor, G.A., and Rivera-Varas, V.V. 2004. Emerging diseases of cultivated potato and their impact on Latin America. Rev Latinoa de La Papa 1:1–8.

Secor, G.A., Rivera, V.V., Abad, J.A., Lee, I., Clover, G.R.G., Liefting, L.W., Li, X., De Boer, S.H. 2009. New diseases and epidemics association of 'Candidatus Liberibacter solanacearum ' with zebra chip disease of potato established by graft and psyllid transmission, electron microscopy and PCR. New Diseases and Epidemics 93:330-335.

Sengoda, V.G., Buchman, J.L., Henne, D.C., Pappu, H.R., and Munyaneza, J.E. 2013. *Candidatus* Liberibacter solanacearum titre over time in *Bactericera cockerelli* (hemiptera: triozidae) after acquisition from infected potato and tomato plants. Journal of Economic Entomology 106:1964–72.

Sengoda, V.G., Cooper, W.R., Swisher, K.D., Henne, D.C., and Munyaneza, J.E. 2014. Latent period and transmission of '*Candidatus* Liberibacter solanacearum' by the potato psyllid *Bactericera cockerelli* (hemiptera: triozidae). PLoS ONE 9:e93475.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G. 2011. Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539.

Sugio, A., MacLean, A.M., Kingdom, H.N., Grieve, V.M., Manimekalai, R., and Hogenhout, S.A. 2011. Diverse targets of pytoplasma effectors from plant development to defense against insects. Annual Reviews of Phytopathology 49:175-195.

Suh, J.R., Herbig, A.K., and Stover, P.J. 2001. New perspectives on folate catabolism. Ann Rev Nutr 21:255–82.

Tamames, J., Moya, A., and Valencia, A. 2007. Modular organization in the reductive evolution of protein-protein interaction networks. Genome Biology 8:94.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725-9.

Tamura, G.S., Kuypers, J.M., Smith, S., Raff, H., and Rubens, C.E. 1994. Adherence of group B streptococci to cultured epithelial cells: roles of environmental factors and bacterial surface components. Infect Immun 62:2450- 2458.

Teresani, G.R., Bertolini, E., Alfaro-Fernández, A., Martínez, C., Tanaka, F.A.O., Kitajima, E.W., Roselló, M., Sanjuan, S., Ferrandiz, J.C., Lopez, M.M., Cambra, M., and Font, M.I. 2014. Association of '*Candidatus* Liberibacter solanacearum' with a vegetative disorder of celery in Spain and development of a real-time PCR method for its detection. Phytopathology 104:804–11.

Teulon, D.A.J., Workman, P.J., Thomas, K.L., Nielsen, M.C. 2009. *Bactericera cockerelli*: incursion, dispersal and current distribution on vegetable crops in New Zealand. New Zealand Plant Protection 62:136–144.

Thompson, S., Fletcher, J.D., Ziebell, H., Beard, S., Panda, P., Jorgensen, N., Fowler, S.V., Liefting, L.W., Berry, L., and Pitman, A.R. 2013. First report of '*Candidatus* Liberibacter europaeus' associated with psyllid infested scotch broom. New Disease Reports 27:5197.

Thompson, S.M., Johnson, C.P., Lu, A.Y., Frampton, R.A., Sullivan, K.L., Fiers, M.W.E.J., Crowhurst, R.N., Pitman, A.R., Scott, I.A.W., Wen, A., Gudmestad, N.C., and Smith, G.R. 2015. Genomes of '*Candidatus* Liberibacter solanacearum' haplotype A from New Zealand and the United States suggest significant genome plasticity in the species. Phytopathology 105:863–871.

Tian, J., Bryk, R., Itoh, M., Suematsu, M., and Nathan, C. 2005. Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of an alpha-ketoglutarate decarboxylase. Proceedings of the National Academy of Sciences of the United States of America 102:10670-10675.

van Ham, R.C.H.J., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J.M., Jimenez, L., Postigo, M., Silva, F.J., Tamames, J., Viguera, E., Latorre, A., Valencia, A., Moran, F., and Moya, A. 2003. Reductive genome evolution in *Buchnera aphidicola*. Proceedings of the National Academy of Sciences of the United States of America 100:581–586.

Wallis, R.L. 1955. Ecological studies on the potato psyllid as a pest of potatoes. USDA Technical Bulletin 1107.

Wann, E.R., Gurusiddappa, S., and Hook, M. 2000. The fibronectin-binding MSCRAMM FnbpA of *Staphylococcus aureus* is a bifunctional protein that also binds to fibrinogen. The Journal of Biological Chemistry 275:13863-13871.

Wang, J., Haapalainen, M., Schott, T., Thompson, S.M., Smith, G.R., Nissinen, A.I., Pirhonen, M. 2017a. Genomic sequence of '*Candidatus* Liberibacter solanacearum' haplotype C and its comparison with haplotype A and B genomes. PLoS ONE 12:e0171531.

Wang, N., Pierson, E.A.., Setubal, J.C., Jin, X., Levy, J., Zhang, Y., Li, J., Rangel, L.T., Martins, J. 2017b. The Candidatus Liberibacter host interface: insights into pathogenesis mechansims and disease control. Annual Review of Phytopathology 55:1.

Winstanley, C., Langille, M.G.I., Fothergill, J.L., Kukavica-Ibrulj, I., Paradis-Bleau, C., Sanschagrin, F., Thomson, N.R., Winsor, G.L., and Quail, M.A. 2009. Newy introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool epidemic strain of Pseudomonas aeruginosa. Genome Research 19:12-23.

Wu, F., Deng, X., Liang, G., Wallis, C., Trumble, J.T., Prager, S., and Chen, J. 2015. *De novo* genome sequence of *Candidatus* Liberibacter solanacearum from a single potato psyllid in California. Genome Announc 3:e01500-15.

Wulff, N.A., Zhang, S., Setubal, J.C., Almeida, N.F., Martins, E.C., Harakava, R., Kumar, D., Rangel, L.T., Foissac, X., Bove, J.M., and Gabriel, D.W. 2014. The complete genome sequence of '*Candidatus* Liberibacter americanus' associated with citrus huanglongbing. Molecular Plant-Microbe Interactions 27:163–176.

Yao, J., Saenkham, P., Levy, J., Ibanez, F., Noroy, C., Mendoza, A., Huot, O., Meyer, D.F., and Tamborindeguy, C. 2016. Interactions Candidatus Liberibacter solanacearum-Bactericera cockerelli: Haplotype effect on vector fitness and gene expression analyses. Front Cell Infect Microbiol 6:62.

Young, P.G., Smith, C.A., Sun, X., Baker, E.N., and Metcalf, P. 2006. Purification, crystallization and preliminary X-ray analysis of Mycobacterium tuberculosis folypolyglutamate synthase. Acta Crystallogr Sec F Struct Biol Cryst Commun 62:579-582.

Zhang, L., Morrison, A.J., and Thibodeau, P.H. 2015. Interdomain contacts and the stability of serralysin protease from *Serratia marcescens.* PLoS ONE 10:e0138419.

Zhou, L., Powell, C.A., Li, W., Irey, M., and Duan, Y. 2013. Prophage mediated dynamics of *Candidatus* Liberibacter asiaticus populations, the destructive bacterial pathogens of citrus Huanglongbing. PLoS ONE 8:e82248.

Zipfel, C., and Robatzek, S. 2010. Pathogen associated molecular pattern triggered immunity: veni, vidi. Plant Physiology 154:551-54.

Zou, H., Gowda, S., Zhou, L., Hajeri, S., Chen, G., and Duan, Y. 2012. The destructive citrus pathogen, '*Candidatus* Liberibacter asiaticus' encodes a functional flagellin characteristic of a pathogen-associated molecular pattern. PLoS ONE 7:e46447.

# Chapter Two

# Changes in *Ca*. L. solanacearum gene expression in plant and insect hosts

Psyllid and plant samples were collected by Annie Barnes and Jessica Verijssen of Plant and Food Research.

## 2.1 Introduction

Bacteria regulate their gene expression to adapt to diverse environments (Yan et al. 2013). In a particular environment, a bacterium will activate a specific gene to provide products to aid its survival within that environment, whereas genes that are not required are not expressed (Chowdhury et al. 1996). In the case of plant pathogenic bacteria, upregulated genes include those required for bacterial adaptation to the host environment during infection (Yan et al. 2013). *Ca*. L. asiaticus *in planta* shows an increase in gene expression of transcriptional regulators, transport systems and secretion systems such as the sec dependent and type 1 secretion system (Yan et al. 2013).

*Ca*. L. solanacearum (*C*Lso) persists in two diverse environments: within the phloem of the plant and as an intracellular bacterium of the psyllid. The phloem of plants contain low levels of essential amino acids while having an abundance of sugars and nutrients (Wernegreen et al. 2002). In order to survive on phloem sap, the psyllid harbours a number of primary endosymbionts such as *Candidatus* Carsonella ruddii and Wolbachia, which are required for the synthesis of essential amino acids for the psyllid. (Fukatsu et al. 1998; Douglas 2006). Thus, the psyllid provides *C*Lso with an increase in amino acid availability along with the presence of fructose, glucose, sucrose and trehalose as carbohydrate metabolites in the psyllid (Killiny and Jones 2018).

*Ca*. Liberibacters persist within an insect vector as well as in the phloem of the plant. The change in environment from *in planta* to within the psyllid host affects regulation in the bacterial pathogen (Yan et al. 2013). Oshima et al. (2011) demonstrated that in *Candidatus* Phytoplasma asteris 33% of genes display differential expression between garland chrysanthemum (*Chrysanthemum coronarium*) and the insect leafhopper (*Macrosteles striifrons*) environments, concluding that genes encoding: transporters, secreted proteins and metabolic enzymes, are regulated in a host-specific manner. Thus, an understanding of the proteins required for pathogen survival within each host could lead to the development of novel methods for disease control.

Similar to *Ca*. L. asiaticus, *C*Lso lacks the common virulence determinants of type II and III secretion systems, thus how the pathogen causes disease and the mechanism for survival

within its hosts remain largely unknown. Furthermore, an understanding of *C*Lso biology is hindered by the unculturable nature of the organism. Thus, there are few studies that investigate the function of enzymes from the organism. Without an understanding of protein function and structure in *C*Lso, there has been little progress towards the identification of an antibacterial target.

## 2.2 Overview of the Chapter

Global gene expression studies allow for the identification of genes that could play a critical role in *C*Lso virulence or are required for survival within the psyllid or plant host. Insight into these gene expression changes suggest key proteins that may be novel targets for antibacterials to treat *C*Lso infections.

The *C*Lso genome (Lin et al. 2011) has indicated several genes that could be implicated in *C*Lso pathogenicity. These genes include *purH*, *secDF* and *ZnMC,* (encoding adenyltransferase/IMP cyclohydrolase, pre-protein translocase and serralysin respectively) could be antibacterial targets if expressed in either the psyllid or *in planta*. This Chapter studies whether *C*Lso gene expression is altered between psyllid and plant hosts. If there is differential gene expression this may aid in determining genes that either play a role in the organisms' survival or are putative pathogenicity determinants.

The absence of both a type II and type III secretion system in the *C*Lso genome has resulted in the organism relying on the type I secretion system and the secretory pathway as the primary mechanisms for the delivery of virulence factors. Serralysin is a bacterial extracellular metalloprotease that is translocated through the type I secretion system and is required for virulence (Ravindran et al. 2018). In *Serratia marcescens*, serralysin inhibits adhesion of immune cells and bacterial clearance (Ishii et al. 2014). *C*Lso retains a serralysin-like gene that Ravindran et al. (2018) speculate the proximity of the gene to divergently transcribed genes encoding the type I secretion system could indicate that the gene is a significant virulence factor for the organism. The Sec pathway has a prominent role in the secretion of major virulence determinants in *Staphylococcus aureus* with deletion of the *secDF* gene resulting in reduced resistance and change in virulence factor expression (Quiblier et al. 2013). *PurH* has been shown to be required for virulence in *Xanthomonas oryzae* pv. *oryzae* (Chatterjee and

Sonti 2005). In addition, *rmlC* and *rmlD* encoding dTDP-4-dehydrorhamnose 3, 5-epimerase and dTDP-4-dehydrorhamnose reductase respectively were also investigated as the proteins encoded by the genes are required for the synthesis of L-rhamnosyl which plays an essential structural role in the cell wall and have been targets for the development of antibacterials against *Mycobacterium tuberculosis* (Ma et al. 2001). *PurH*, *ZnMC*, *secDF*, *rmlC* and *rmlD* are known to be required for virulence in a number of organsisms, thus these genes could play a key role in *C*Lso pathogenicity. Here, the differential expression of *purH*, *secDF*, *ZnMC*, *rmlC* and *rmlD* were investigated in order to determine whether the genes were upregulated *in planta* which could suggest the genes were required for *C*Lso disease establishment.

This study also aimed to understand whether there were essential genes required for the survival within the psyllid or plant host that could present novel targets for antibacterial design. *C*Lso is unable to synthesise histidine, tryptophan, and thiamine and has reduced ability to produce phenylalanine and tyrosine compared *Liberibacter crescens* (Leonard et al 2012). Therefore it is believed that the remaining amino acid biosynthesis pathways must be important for the survival of *C*Lso either within the psyllid or in planta. The *coaBC*, *dapA* and *argJ* encoding phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase, dihydrodipicolinate synthase and glutamate N-acetyltransferase, respectively, were also analysed in this study. Phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase (encoded by *coaBC*) is required for L-cysteine synthesis, whilst dihydrodidpicolinate synthase (encoded by *dapA*) is necessary in L-lysine synthesis and *argJ* for L-arginine production. In addition, *pykF*, encoding pyruvate kinase was also investigated for differential expression in the psyllid and plant as it is essential for glycolysis and *C*Lso survival.

This gene expression study provides insight into key genes that could play a role in pathogenicity and survival of *C*Lso within the psyllid. Thus, with an understanding of the changes of *C*Lso gene expression in psyllids and in plants, a rational approach for determining promising proteins for future protein expression and characterisation studies could be developed.

## 2.3 Results and Discussion

To compare the gene expression of putative pathogenicity and essential survival genes of *C*Lso in its plant host and its insect host psyllid, the quantification of gene expression via quantitative reverse transcription-polymerase chain reaction (qRT-PCR) was conducted. Quantitative reverse transcription-polymerase chain reaction is used to measure changes in gene expression. Reverse transcriptase uses mRNA as a template to produce cDNA which is used as a template for the qRT-PCR reaction (Valasek and Repa 2005). The amplification of template is monitored by fluorescence (Freenan et al. 1999).

Primers specific for *dapA, znMC, secDF, purH, pykF, rmlC, rmlD, coaBC and argJ* were designed (Chapter 8, methods). Primers were chosen based on the specificity of their amplification and tested using DNA extracted from *C*Lso infected and uninfected plant and psyllid samples. In order to effectively determine which genes were differentially expressed, all genes were normalized against the housekeeping gene *recA* (recombinase A). The *recA* gene was chosen from a panel of housekeeping genes (Chapter Eight, Table 8.18, Section 8.11.1) for its low levels of variation in gene expression between *C*Lso infected and uninfected plant and psyllid RNA samples. Fold change in *C*Lso gene expression levels was determined via the use of the double delta -Ct analysis (Chapter Eight, Section 8.11.6).

### 2.3.1 Changes in *C*Lso gene expression in psyllid and *in planta*

A Student's *t*-test was used to determine statistically significant changes in *C*Lso gene expression between infected psyllids and plants. D*apA*, *purH* and *rmlD* showed increased expression in the *C*Lso psyllid host, the *t*-test shows that both genes have P values of less than 0.01 which indicates that both genes were upregulated in psyllids. In contrast, *secDF* and *ZnMC* genes had higher expression levels *in planta*, the *t*-test shows the data has a P value of less than 0.01 (Figure 2.1). It should be noted that *t*-test analysis determined that *C*Lso *pykF* was upregulated in infected plants with a P value less than 0.05.

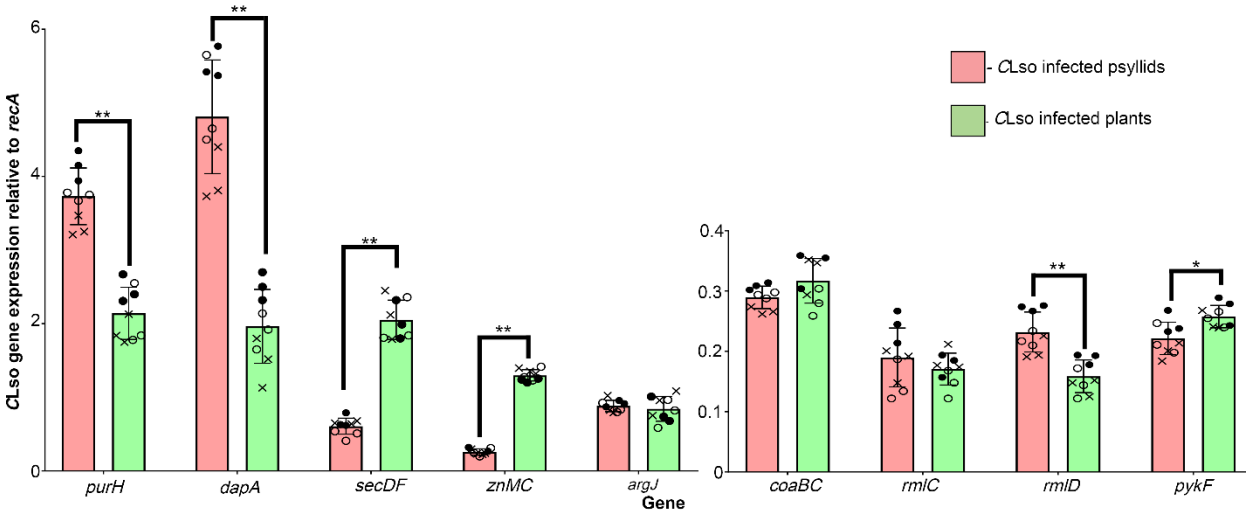



**Figure 2.1**: Gene expression of *C*Lso genes in psyllid and in planta based on qRT-PCR. Mean ± standard deviation for biological (o, ●, ×) and three technical replicates. Asterisks indicate statistically significant differences in expression levels between *C*Lso infected psyllids and plants (*P ≤ 0.05, **P ≤ 0.01; *t*-test).

The *purH*, *dapA* and *rmlD* genes show a fold change in gene expression of 0.73, 1.45 and 0.15 respectively, in the psyllid. This indicates that *purH*, *dapA* and *rmlD* are upregulated in psyllids (Table 2.1). The fold change for *secDF* and *ZnMC* was -0.71 and -0.84 indicating that these genes were upregulated *in planta* (Table 2.1). *C*Lso *pykF* shows a fold change in gene expression of -0.14 *in planta* compared to the psyllid, which demonstrates the gene is upregulated in infected plants.

**Table 2.1:** Expression profile of genes of *Candidatus* Liberibacter solanacearum in psyllid compared to *in planta.*

| Accession Number[1] | Gene | Gene Product | Fold Change[2] | P-value |
|---|---|---|---|---|
| KJZ81084.1 | *purH* | adenyltransferase/IMP cyclohydrolase | 0.73 ± 0.75 | 0.0039 |
| KJZ81864.1 | *dapA* | dihydrodipicolinate synthase | 1.45 ± 1.15 | 0.0039 |
| WP_034441786.1[3] | *secDF* | pre protein translocase-SECDF | -0.71 ± 0.4 | 0.0039 |
| WP_034442221.1[3] | *ZnMC* | zinc dependent metalloprotease-serralysin | -0.84 ± 0.55 | 0.0039 |
| KJZ81994.1 | *coaBC* | phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase | 0.02 ± 0.18 | 0.5703 |
| KJZ82379 | *argJ* | glutamate N-acetyltransferase | -0.044 ± 2.16 | 0.5703 |
| KJZ81768.1 | *rmlC* | dTDP-4-dehydrorhamnose 3,5-epimerase | -0.03 ± 0.45 | 0.5703 |
| KJZ81766.1 | *rmlD* | dTDP-4-dehydrorhamnose reductase | 0.15 ± 0.38 | 0.0039 |
| WP_034441313.1[3] | *pykF* | pyruvate kinase | -0.14 ± 0.42 | 0.0195 |

[1] The accession number provided is for the *C*Lso NZ1 genome

[2] Fold change ($log_2$) is the relative gene expression in psyllid versus in planta for *C*Lso. Positive values indicate the gene was overexpressed in psyllids compared with *in planta* whilst negative values indicate the gene was overexpressed *in planta* compared with in the psyllid.

[3]Referenced from NZ_JMTK01000002 (*C*Lso-NZ-1 complete genome, scaffold 2)

One aim of the research was to determine whether there was a differential expression between the psyllid and in planta of putative pathogenicity genes. Upregulation of genes *in planta* could indicate their role in host colonisation and disease establishment. *C*Lso lacks the common virulence determinants of a type II and III secretion, but retains both a type I and a secretory pathway (Sec). Upregulation *in planta* of both *ZnMC*, which encodes a putative serralysin like protein and is a type I secretory system substrate, and *secDF* (part of the Sec machinery) is consistent with the theory that both proteins encoded by the genes could be essential for *C*Lso pathogenicity (Lin et al. 2011; Prasad et al. 2016; Ravindran et al. 2016).

Furthermore, with their potential role in *C*Lso pathogenicity, *ZnMC* and *secDF* are potential candidates for further protein functional characterisation.

Interestingly, the protein encoded by *purH* (adenyltransferase/IMP cyclohydrolase) has been implicated in virulence in *S. pneumoniae* (Polissi et al. 1998), *Vibrio vulnificus* (Kim et al. 2003), *Salmonella enterica* serovar Typhimurium (McFarland and Stocker 1987) and *Bacillus anthracis* (Jenkins et al. 2011). Here, *C*Lso *purH* was expressed in both infected psyllid and *in planta*; however, there were higher expression levels of the gene in the psyllid. Ramsey et al. (2010) show that the pea aphid, *Acyrthosiphon pisum*, lacks genes required for purine synthesis. The aphid's requirement for purine is provided by complementarity of the purine metabolic capabilities of its symbiotic bacterium *Buchnera aphidicola* (Ramsey et al. 2010). In addition, Yao et al. (2016) propose that differential expression of *purH* within the *C*Lso haplotypes A and B could account for the differences in pathogenicity, although, within the psyllid there was no change in *purH* gene expression between the haplotypes. However, Yao et al. (2016) did not investigate the differential expression of *purH in planta*. Here, a comparison between the gene *in planta* and in psyllid is presented with *purH* upregulated in the psyllid host. One hypothesis is that, like *B. aphidicola* and the *A. pisum* (pea aphid) (Wernegreen 2002; Wu et al. 2006), *C*Lso supplements purine biosynthesis for the psyllid. In the absence of the *B. cockerelli* psyllid genome for analysis of whether the psyllid contains genes required for purine biosynthesis, this remains speculative. The high expression levels of *purH* in infected plants and psyllids could indicate that the gene is required for the organism to survive in both hosts. Thus, *purH* could be an antimicrobial target. An understanding of the functionality and structure of the adenyltransferase/IMP cyclohydrolase enzyme encoded by *purH* would be required to further investigate the essentiality of the gene for *C*Lso.

Of the genes required for amino acid biosynthesis, only *dapA* was upregulated in psyllids. This indicates that *dapA* could be needed for survival within the psyllid host. A possible advantage of the *C*Lso and psyllid endosymbiotic relationship is the ability for the eukaryote host to acquire metabolic capabilities from unique ecological environments (Wu et al. 2006). In the case of *B. aphidicola*, the bacterium allows the pea aphid to feed on sap from plant phloem (Wu et al. 2006). Despite the presence of some nutrients, phloem has few essential amino acids that are unable to be synthesized by insects. In this case, *Buchnera* complements amino

acid production for the aphids (Wernegreen 2002; Wu et al. 2006). A similar complementary relationship could occur between *C*Lso and the psyllid.

Rhamnose is a sugar found widely in bacteria and plants. The *rmlD* gene results in colonisation defect in *Vibrio cholera* mutants (Chiang and Mekalanos 1999) and inhibiton of cell wall polysaccharide synthesis in *Streptococcus mutans* (Yamashita et al. 1999). As *rmlD* is believed to play a role in bacterial pathogenicity, it was surprising that an upregulation of the gene was observed in the psyllid. However, it is possible that an upregulation of *rmlD* is due to a lack of the rhamnose sugar in the psyllid. Thus, *C*Lso utilises rhamnose produced by the plant which could account for the low levels of *rmlC* and *rmlD* gene expression.

Interestingly, *Ca*. L. asiaticus shows an increase in gene expression *in planta* of eight glycolysis related proteins: glucokinase, fructose-bisphosphate aldolase, phosphoglycerate kinase, glyceraldehyde 3-phospahte dehydrogenase, phosphopyruvate hydratase, dihydrolipoamide dehydrogenase, phosphoenolpyruvate carboxykinase and phosphoglucomutase) (Yan et al. 2013). Although *pykF* in *Ca*. L. asiaticus was not upregulated in plants, the high expression levels of glycolysis related genes *in planta* indicate that the organism is able to use glucose acquired from the host plant as an energy source to support intracellular growth of the organism within the plant. Further, glycolysis is upregulated during infection and is required for virulence in *Yersisnia pseudotuberculosis* (Chaudhuri et al. 2009) and *Salmonella enterica* serovar *Typhimurium* (Rosso et al. 2008). Pyruvate kinase, encoded by *pykF*, is required for the conversion of phosphoenolpyruvate to pyruvate, upregulation of *C*Lso *pykF in planta* suggests the gene could be of interest in order to understand *C*Lso metabolism and as a novel antibacterial target.

The aim of this gene expression study was to identify genes that may be involved in *C*Lso pathogenicity or required for survival within either the psyllid or plant hosts, which in turn could provide insight into probable proteins for future functional characterisation and further inhibitor development. Here, the *secDF* and *ZnMC* genes were upregulated *in planta,* which is consistent with the hypothesis that the genes could be required for pathogenicity (Lin et al. 2011; Wang et al. 2016). Additionally, the upregulation of *pykF* could be associated with the requirement of glycolysis for the growth of *C*Lso within the plant. Contrastingly, *purH* and *dapA* are upregulated in psyllids. It is proposed that these genes could be upregulated in order

to complement loss of function of purine and lysine biosynthesis in the psyllid making the genes ideal antimicrobial targets. Thus, *secDF, ZnMC, pykF, purH* and *dapA* were selected to further explore their function within *C*Lso. Further insights into the function and structure of *purH, pykF* and *dapA* are discussed in this thesis. Serralysin encoded by *ZnMC* was probed in Chapter Four in relation to expression and purification of the protein, whilst further investigation of the *secDF* membrane protein was not undertaken as membrane proteins are notoriously difficult to isolate *in vitro,* in addition, previous studies determined the difficulty in expressing soluble proteins from *C*Lso (Ravindran et al. 2019)

## 2.4 Summary

This Chapter, along with previous gene expression analysis, (Nachappa et al. 2012; Levy et al. 2017) have provide insight into genes that could play a role in *C*Lso pathogenicity and survival, however, there is still a need to validate the structure and function of the annotated genes. Furthermore, antibacterial design specific to *C*Lso relies on an understanding of structure, function and inhibition, which currently remains absent for any protein in the *C*Lso genome. This Chapter reports an investigation into the changes in *C*Lso gene expression in psyllids and *in planta*. The work provides five genes*: dapA, purH, pykF,, secDF* and *ZnMC* as candidates for protein functional and structural characterisation based on their putative role in pathogenicity (*secDF* and *ZnMC*) and proposed role as genes required to complement the loss of a biosynthetic pathway within the psyllid (*dapA* and *purH*) and plant (*pykF*) hosts.

## 2.5 References

Chatterjee, S., and Sonti, R.V. 2005. Virulence deficiency caused by a transposon insertion in the *purH* gene of *Xanthomonas oryzae* pv. *oryzae*. Canadian Journal of Microbiology 51(7):575-81.

Chaudhuri, R.R., Sebaihia, M., Hobman, J.L., Webber, M.A., Leyton, D.L., Goldberg, M.D., Cunningham, A.F., Scott-Tucker, A., Ferguson, P.R., Thomas, C.M., and Frankel, G. 2010. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli strain* 042. PloS One 5(1):e8801.

Chiang, S.L., and Mekalanos, J.J. 1999. rfb mutations in *Vibrio cholerae* do not affect surface production of toxin-coregulated pili but still inhibit intestinal colonization. Infection and Immunity 67(2):976-80.

Chowdhury, R., Sahu, G.K., and Das, J. 1996. Stress response in pathogenic bacteria. Journal of Biosciences 21(2):149-60.

Douglas, A.E. 2006. Phloem-sap feeding by animals: problems and solutions. Journal of Experimental Botany 57(4):747-54.

Freeman, W.M., Walker, S.J., and Vrana, K.E. 1999. Quantitative RT-PCR: pitfalls and potential. Biotechniques 26(1):112-25.

Fukatsu, T., and Nikoh, N. 1998. Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). Applied Environmental Microbiology 64(10):3599-606.

Ishii, K., Adachi, T., Hara, T., Hamamoto, H., and Sekimizu, K. 2014. Identification of a *Serratia marcescens* virulence factor that promotes hemolymph bleeding in the silkworm, *Bombyx mori*. Journal of Invertebrate Pathology 117:61-7.

Jenkins, A., Cote, C., Twenhafel, N., Merkel, T., Bozue, J., and Welkos, S. 2011. Role of purine biosynthesis in *Bacillus anthracis* pathogenesis and virulence. Infection and Immunity 79(1):153-66.

Killiny, N., and Jones, S.E. 2018. Metabolic alterations in the nymphal instars of *Diaphorina citri* induced by *Candidatus* Liberibacter asiaticus, the putative pathogen of huanglongbing. PloS One 13(1):e0191871.

Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, E.K., Shin, D.H., Chung, S.S., Choy, H.E., Progulske-Fox, A., Hillman, J.D., and Handfield, M. 2003. Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. Infection and Immunity 71(10):5461-71.

Leonard, M.T., Fagen, J.R., Davis-Richardson, A.G., Davis, M.J., and Triplett, E.W. 2012. Complete genome sequence of *Liberibacter crescens* BT-1. Standards in Genomic Sciences 7(2):271.

Levy, J.G., Mendoza, A., Miller, J.C., Tamborindeguy, C., and Pierson, E.A. 2017. Global gene expression in two potato cultivars in response to '*Candidatus* Liberibacter solanacearum' infection. BMC Genomics 18(1):960.

Lin, H., Han, C.S., Liu, B., Lou, B., Bai, X., Deng, C., Civerolo, E.L., and Gupta, G. 2013. Complete genome sequence of a Chinese strain of "*Candidatus* Liberibacter asiaticus". Genome Announcements 1(2):e00184-13.

Ma, Y., Stern, R.J., Scherman, M.S., Vissa, V.D., Yan, W., Jones, V.C., Zhang, F., Franzblau, S.G., Lewis, W.H., McNeil, and M.R. 2001. Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose. Antimicrobial Agents and Chemotherapy 45(5):1407-16.

McFarland, W.C., and Stocker, B.A. 1987. Effect of different purine auxotrophic mutations on mouse-virulence of a Vi-positive strain of *Salmonella dublin* and of two strains of *Salmonella typhimurium*. Microbial Pathogenesis 3(2):129-41.

Nachappa, P., Shapiro, A.A., and Tamborindeguy, C. 2012. Effect of '*Candidatus* Liberibacter solanacearum' on fitness of its insect vector, *Bactericera cockerelli* (Hemiptera: Triozidae), on tomato. Phytopathology 102(1):41-6.

Oshima, K., Kakizawa, S., Arashida, R., Ishii, Y., Hoshi, A., Hayashi, Y., Kagiwada, S., and Namba, S. 2007. Presence of two glycolytic gene clusters in a severe pathogenic line of *Candidatus* Phytoplasma asteris. Molecular Plant Pathology 8(4):481-9.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. 1998. Large-scale identification of virulence genes from *Streptococcus pneumoniae*. Infection and Immunity 66(12):5620-9.

Prasad, S., Xu, J., Zhang, Y., and Wang, N. 2016. SEC-translocon dependent extracytoplasmic proteins of *Candidatus* Liberibacter asiaticus. Frontiers in Microbiology 20(7):1989.

Quiblier, C., Seidl, K., Roschitzki, B., Zinkernagel, A.S., Berger-Bächi, B., and Senn, M.M. 2013. Secretome analysis defines the major role of SecDF in *Staphylococcus aureus* virulence. PLoS One 8(5):e63513.

Ramsey, J.S., MacDonald, S.J., Jander, G., Nakabachi, A., Thomas, G.H., and Douglas, A.E. 2010. Genomic evidence for complementary purine metabolism in the pea aphid, *Acyrthosiphon pisum*, and its symbiotic bacterium *Buchnera aphidicola*. Insect Molecular Biology 19:241-8.

Ravindran, A., Saenkham, P., Gad-Levy, J., Tamborindeguy, C., Lin, H., Gross, D., Pierson, E.A. 2018. Characterization of the serralysin-like gene of *Candidatus* Liberibacter solanacearum associated with potato zebra chip disease. Phytopathology 108(3): 327–335.

Rosso, M.L., Chauvaux, S., Dessein, R., Laurans, C., Frangeul, L., Lacroix, C., Schiavo, A., Dillies, M.A., Foulon, J., Coppée, J.Y., and Médigue, C. 2008. Growth of *Yersinia pseudotuberculosis* in human plasma: impacts on virulence and metabolic gene expression. BMC Microbiology 8(1):211.

Valasek, M.A., and Repa, J.J. 2005. The power of real-time PCR. Advances in Physiology Education 29(3):151-9.

Wernegreen, J.J. 2002. Genome evolution in bacterial endosymbionts of insects. Nature Reviews Genetics 3(11):850.

Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., and Moran, N.A. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biology 4(6):e188.

Yamashita, Y., Shibata, Y., Nakano, Y., Tsuda, H., Kido, N., Ohta, M., and Koga, T. 1999. A novel gene required for rhamnose-glucose polysaccharide synthesis in *Streptococcus mutans*. Journal of Bacteriology 181(20):6556-9.

Yan, Q., Sreedharan, A., Wei, S., Wang, J., Pelz-Stelinski, K., Folimonova, S., and Wang, N. 2013. Global gene expression changes in *Candidatus* Liberibacter asiaticus during the transmission in distinct hosts between plant and insect. Molecular Plant Pathology 14(4):391-404.

Yao, J., Saenkham, P., Levy, J., Ibanez, F., Noroy, C., Mendoza, A., Huot, O., Meyer, D.F., and Tamborindeguy, C. 2016. Interactions "*Candidatus* Liberibacter solanacearum"—*Bactericera cockerelli*: haplotype effect on vector fitness and gene expression analyses. Frontiers in Cellular and Infection Microbiology 9(6):62.

# Chapter Three

## Biophysical characterisation and structural insights into the *Ca*. L. solanacearum adenyltransferase/IMP cyclohydrolase

## 3.1 Introduction

In chapter two, gene expression analyses shows upregulation of *purH* relative to the housekeeping gene, *recA*, in psyllids compared to *in planta* and high expression levels overall which could indicate the requirement of purine synthesis for survival in both the psyllid and within the plant. *PurH* encodes the adenyltransferase/IMP cyclohydrolase and is required for virulence in *S. pneumoniae* (Polissi et al. 1998), *Vibrio vulnificus* (Kim et al. 2003), *Salmonella enterica* serovar Typhimurium (McFarland and Stocker 1987) and *Bacillus anthracis* (Jenkins et al. 2011). Thus, an upregulation of the *purH* could indicate that ATIC could be required for *C*Lso pathogenicity. Thus, in order to determine whether the ATIC plays a role in pathogen survival in both hosts and/or pathogenicity, a further understanding of the enzyme function, structure and catalysis is required. Here, a brief introduction to *de novo* purine synthesis in *C*Lso is discussed. In addition, the ATIC catalytic mechanism and structure are also introduced as this knowledge is required for understanding variations in *C*Lso ATIC.

### 3.1.1 Purine synthesis and the Liberibacter genome

Purine nucleotides are essential for DNA and RNA production (Le Nours et al. 2011), thus the *de novo* purine biosynthetic pathway is critical for growth and survival (Sassetti et al. 2003). Purine nucleotide generation can be *via* the *de novo* purine biosynthesis pathway or the salvage pathway. *De novo* purine synthesis generates inosine monophosphate from a series of reactions with phosphoribosyl pyrophosphate as the initial substrate (Liechti and Goldberg 2012) and is fed by the metabolism of various amino acids (Liechti and Goldberg 2012). Alternatively, pre-existing purine bases and nucleotides are salvaged from the surrounding environment and added directly to phosphoribosyl pyrophosphate to generate nucleotides that are subsequently utilised in the nucleotide biosynthesis pathway to generate GMP or AMP (Becerra and Lazcano 1996; Ramsey et al. 2010). The latter process is more energetically favourable.

Interestingly, both *Ca*. L. solanacearum (*C*Lso) and *Ca*. L. asiaticus appear to have lost the genes required for purine salvage (Duan et al. 2009; Lin et al. 2011). Liberibacters maintain genes required for *de novo* purine synthesis with the pentose phosphate pathway providing a potential supply of 5-phosphoribosyl disphosphate to feed into purine synthesis. However,

Hartung et al. (2011) noted that *Ca*. L. asiaticus is deficient in its ability to synthesise purines as it only retains 47 annotated genes for purine metabolism compared to the 95 purine metabolic genes annotated for *Sinorhizobium meliloti*. Thus, it was hypothesized that *Ca*. L. asiaticus was able to produce nucleotides from nucleoside disphosphates, but unable to synthesise guanine, xanthine, hypoxanthine or adenine (Hartung et al. 2011). Furthermore, Duan et al. (2009) and Lin et al. (2011) postulate that both *C*Lso and *Ca*. L. asiaticus have lost the ability to synthesise and metabolise purines, thus they must be obtained from the host as di- or tri-nucleosides. The loss of purine synthesis is consistent with the adaptation to an intracellular lifestyle, whereby deletion of redundant genes further contributes to genome reduction (Toft and Andersson 2010). The lack of annotated purine salvage orthologous genes does not necessarily mean *C*Lso is incapable of purine salvage (Duan et al. 2009) as relaxed substrate specificity of enzymes has compensated for loss of function in the reduced genome organism *Mycoplasma* species (Glass et al. 2006).

### 3.1.2 Challenges in understanding *C*Lso protein function

Although *C*Lso retains genes for purine biosynthesis, there is still a question as to whether the annotated genes are functional. Chapter One, Section 1.6 briefly describes the problems associated with genome wide annotations and the occurrence of errors in these annotations. Therefore, although bioinformatic approaches are capable of annotation, there is still a need to relate the biochemical and cellular function to the uncharacterised gene. In one example, the annotated *menG* gene from *Mycobacterium tuberculosis* was predicted to encode an S-adenosylmethionine-dependent methyltransferase, however extensive protein characterisation and structural analysis determined the encoded protein has no structural similarity to the suggested annotation (Johnston et al. 2003). This highlights the importance of gene validation through protein characterization in order to relate annotated gene to protein function for *Ca*. L. solanacearum. In addition, an understanding of *C*Lso protein structure and function is limited. The majority of the attempts to obtain *C*Lso proteins utilised an approach requiring the denaturation and refolding of the proteins. Ravindran et al. (2018) purified the serralysin-like protein (haplotype LsoA KJZ82330.1 and haplotype LsoB WP_013461860 LsoA) from *Candidatus* Liberibacter solanacearum haplotypes LsoA and LsoB (HenneA and ZC1) under denaturing conditions. However, the purified enzyme had no

measurable protease activity, leading to the conclusion that either the putative gene does not encode a functional enzyme, or that the protein incorrectly folds during the purification and refolding process (Ravindran et al. 2018). This approach has also been applied for the expression of *Ca.* L. asiaticus proteins (Lau et al. 2011; Liu et al. 2017), without subsequent protein characterisation. In general, the strategy has been to denature and then refold *Candidatus* Liberibacter proteins—a harsh process that can lead to misfolding or aggregation (Fink 1998). Therefore, proteins solubilised by this method should be extensively characterised to ensure loss of function is a feature of the enzyme rather than a consequence of the refolding method.

### 3.1.3 The reaction catalysed by ATIC

In order to determine whether an enzyme is functional, an understanding of the reaction the enzyme catalyses is required. Here, a brief introduction to the reaction ATIC catalyses is presented. The catalytic mechanism is also presented and highlights key residues required for ATIC activity.

The *de novo* purine biosynthetic pathway ultimately converts 5-phosphoribosyl disphosphate to inosine monophosphate (Greasley et al. 2001; Bulock et al. 2002; Le Nours et al. 2011). In bacteria and eukaryotes, ATIC contains two domains that catalyse the penultimate and final steps in the *de novo* purine biosynthesis pathway (Greasley et al. 2001; Vergis et al. 2001; Bulock et al. 2002; Cheong et al. 2004). However, in some Archaea folate is not required for the adenyltransferase/IMP cyclohydrolase activity (White 1997). Genomic DNA sequence comparisons indicate the absence of *purH* genes in the domain Archaea (Koike et al. 1999). The 5-amino-4-imidazolecarboxamide ribonucleotide formyl transferase domain (AICARFT), located in the C-terminal region, catalyses the transfer of the formyl group from a reduced folate cofactor, $N^{10}$-formyltetrahydrofolate to the exocyclic amino group of the purine precursor 5-amino-4-imidazolecarboxamide ribonucleotide (AICAR) to form 5-formylaminoimidazole-4-carboxamide ribonucleotide (FAICAR) (Greasley et al. 2001; Le Nours et al. 2002; Wolan et al. 2002; Wolan et al. 2004). The inosine monophosphate cyclohydrolase domain (IMPCH) resides in the N-terminal region and catalyses the subsequent ring closure reaction to form inosine monophosphate and water (Greasley et al. 2001; Le Nours et al. 2002; Wolan et al. 2002; Wolan et al. 2004) (Figure 3.1).

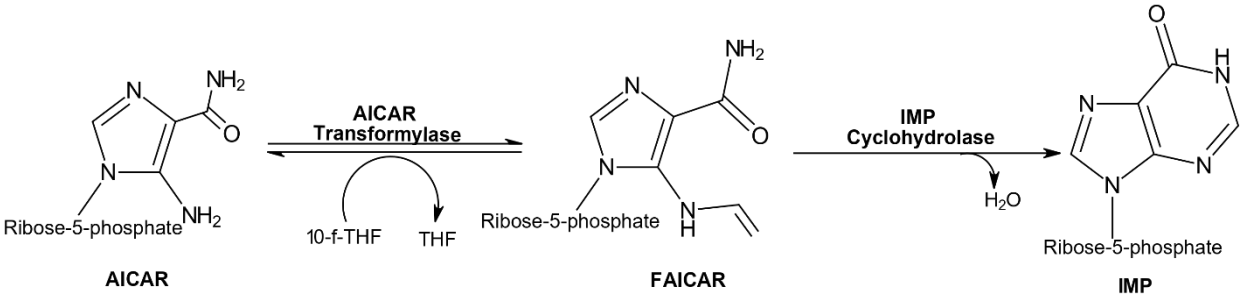



**Figure 3.1:** The reaction catalysed by adenyltransferase/IMP cyclohydrolase.

*3.1.3.1 The catalytic mechanism of ATIC*

The avian (Greasley et al. 2001; Wolan et al. 2002) and human (Wolan et al. 2004) ATIC enzymes have been extensively structurally characterized. In addition, kinetic, mutagenesis and inhibition studies have provided an understanding of the catalytic mechanisms for each reaction (Wolan et al. 2003; Vergis et al. 2004).

Analysis of the crystal structures of avian (Wolan et al. 2002; Wolan et al. 2003), human (Greasley et al. 2001), *Mycobacterium tuberculosis* (Le Nours et al. 2011) and *Thermotoga maritima* ATIC (Axelrod et al. 2008) showed that the enzyme exists as an extensively intertwined homodimer, with dimerization required for catalytic activity. The 5-amino-4-imidazolecarboxamide ribonucleotide transferase and inosine monophosphate cyclohydrolase activities reside in different domains of the polypeptide; these domains have been expressed individually from human ATIC and are enzymatically active (Rayl et al. 1996).

Beardsley et al. (1998) and Bulock et al. (2002) probed the kinetic mechanism of human bifunctional ATIC and found that the folate cofactor binds first at the 5-amino-4-imidazolecarboxamide ribonucleotide transferase active site. This formyl group is then transferred from 10-formyl-tetrahydrofolate to 5-amino-4-imidazolecarboxamide ribonucleotide. The His267 residue of the 5-amino-4-imidazolecarboxamide ribonucleotide transferase domain increases the nucleophilicity of the 5-amino group of 5-amino-4-imidazolecarboxamide ribonucleotide and acts as a catalytic base to deprotonate the 5-amino-4-imidazolecarboxamide ribonucleotide amino group followed by nucleophilic attack of the substrate formyl group on 10-formyl-tetrahydrofolate (Wolan et al. 2002). Lys266

stabilises the oxyanion transition state and carries protons to the $N_{10}$ of tetrahydrofolate (Figure 3.2).

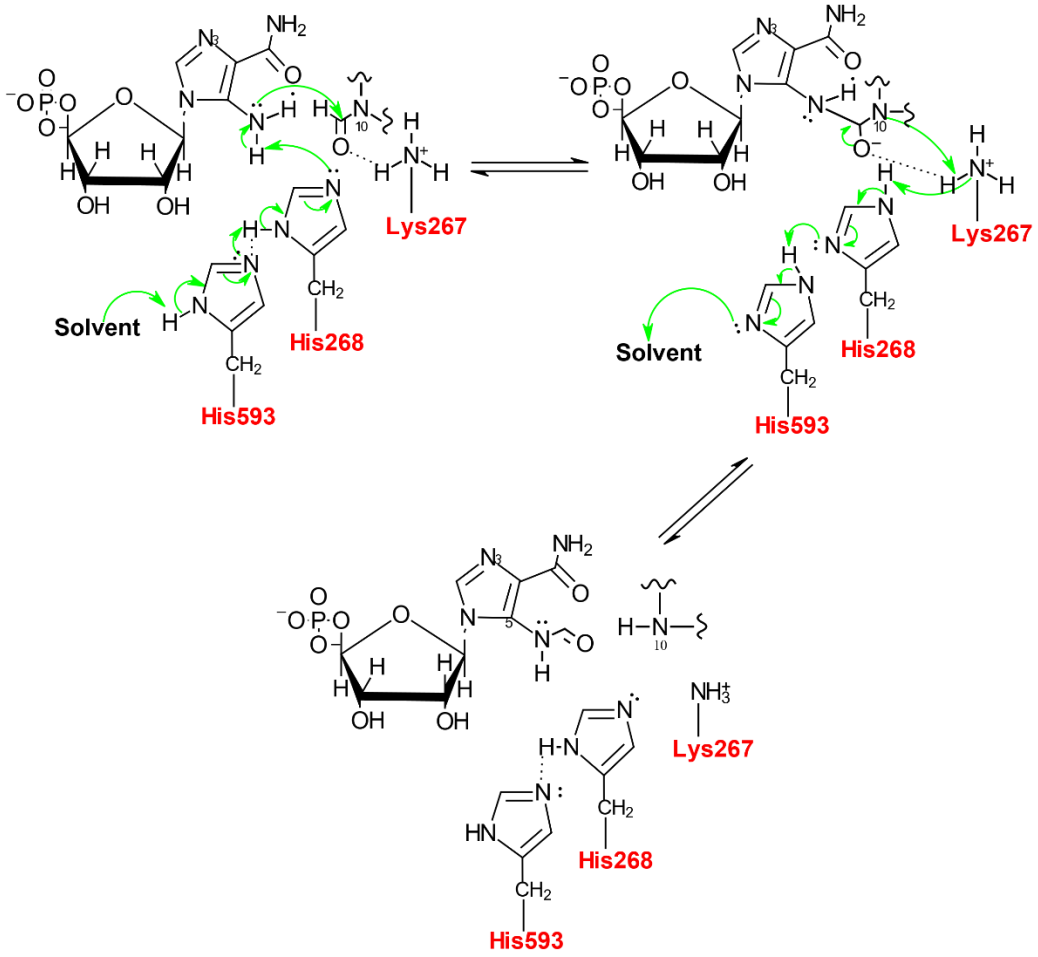



**Figure 3.2:** The mechanism for formyl transfer in the 5-amino-4-imidazolecarboxamide ribonucleotide transferase domain. The removal of a proton from the 5-amino group of 5-amino-4-imidazolecarboxamide ribonucleotide during formyl transfer and the nucleophilic attack on 10-formyltetrahydrofolate are shown (green arrows). The substrate protein is shuttled out by an interaction between His593 (shown in red as the residue is on the other monomer) and the bulk solvent leaving Lys267 to protonate the N10 group of tetrahydrofolate through proton exchange with the solvent or with His268 (adapted from Wolan et al. 2002).

The inosine monophosphate cyclohydrolase active site of both human and avian ATIC has been proposed to favourably orient 5-formylaminoimidazole-4-carboxamide ribonucleotide (4-carboxamide and the 5-formylamine) to promote intramolecular nucleophilic attack, subsequent cyclization and elimination of water from the reaction intermediate (Wolan et al. 2004). This reorientation is facilitated by a hydrogen bond donor, Thr67, and a main chain

hydrogen bond acceptor, Arg64. When binding 5-formylaminoimidazole-4-carboxamide ribonucleotide in the appropriate conformation, there is a conformational rearrangement that allows the side chain of Tyr104 to flip into the active site to sequester the substrate (Wolan et al. 2004) (Figure 3.3).

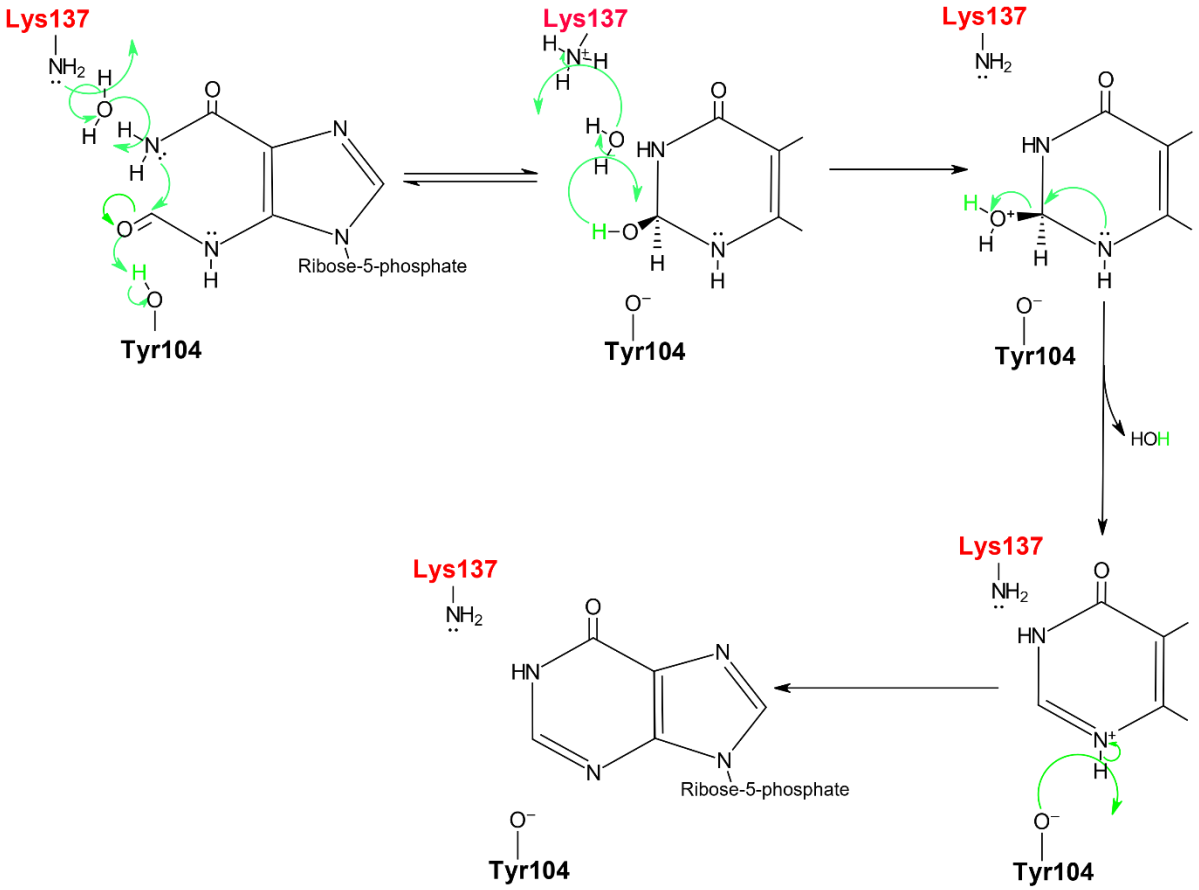



**Figure 3.3:** Mechanism for the inosine monophosphate cyclohydrolase reaction of adenyltransferase/IMP cyclohydrolase. Lys137 is located on the other monomer (coloured red) of the protein and therefore the reaction is facilitated by the dimerisation of the enzyme (adapted from Vergis and Beardsley 2004).

### 3.1.4 The structure of ATIC

The function of an enzyme is reliant on the structure of a protein. The Chapter aims to structurally characterize *C*Lso, as a result the structure of ATIC is introduced.

Domain structures for human, avian, *T. maritima* and *M. tuberculosis* (Greasley et al. 2001; Wolan et al. 2002; Wolan et al. 2003; Axelrod et al. 2008; Le Nours et al. 2011) have characterized ATIC as a dimer with two domains joined by a linker β-ribbon (Le Nours et al.

2011). The inosine monophosphate cyclohydrolase domain has a similar structure in all organisms, with a modified Rossmann fold topology and a central 5-stranded β-sheet with a strand order 3-2-1-4-5. Six helices are packed around the fold with a long helix that leads into the linker region. The active site, which cyclizes 5-formylaminoimidazole-4-carboxamide ribonucleotide to inosine monophosphate, occurs along the β2-α2 loop across strands β1 and β4 to a site between helix α8 and a loop joining residues 71–74 (Le Nours et al. 2011) (Figure 3.4).

The C-terminal amino-4-imidazolecarboxamide ribonucleotide transferase domain consists of two tandem subdomains, both of which are composed of a α + β fold of a mixed six-stranded β sheet and four α helices (Greasley et al. 2001; Wolan et al. 2002; Wolan et al. 2003; Axelrod et al. 2008; Le Nours et al. 2011). The formyl transfer active site is situated at the interface between the two monomers of the dimer, forming two active sites per dimer. Avian and human ATIC structures with folate analogues and bisubstrate inhibitors indicated that one monomer predominantly binds AICAR whilst the other binds folate, indicating that dimerisation must be essential for formyl transfer (Greasley et al. 2001; Wolan et al. 2002; Wolan et al. 2003) (Figure 3.4).

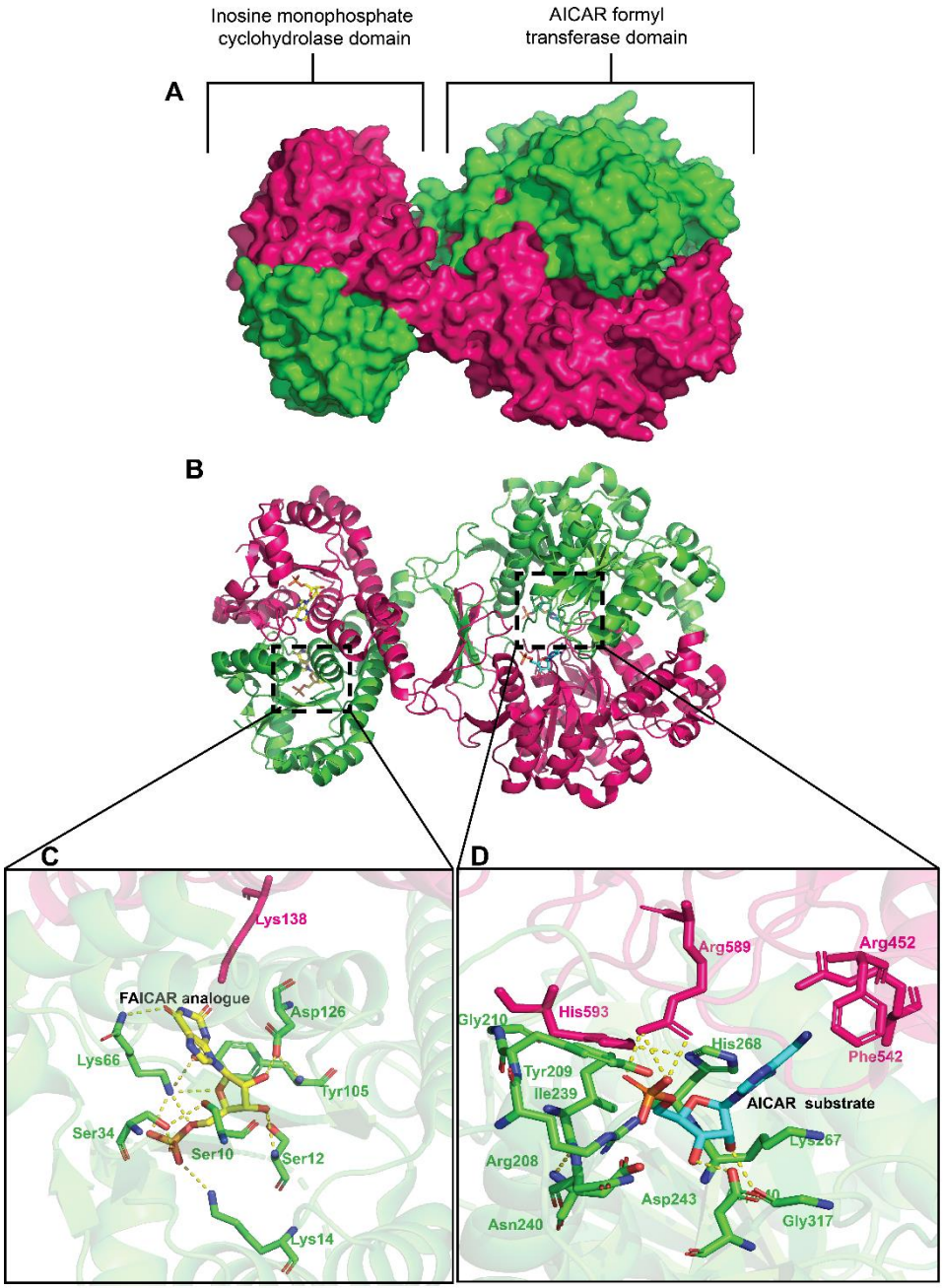



**Figure 3.4:** Structure of the avian adenyltransferase/IMP cyclohydrolase (PDB ID: 9M8N). **A** is the space filled dimer of the enzyme with each monomer coloured pink and green, while **B** is the dimer in a cartoon model. **C** shows an analogue (xanthosine 5'-monophosphate) of 5-formamidoimidazole-4-carboxamide ribotide in the active site of the inosine monophosphate cyclohydrolase domain and shows the residues required for catalytic activity. **D** shows 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) in the active site of the amino-4-imidazolecarboxamide ribonucleotide transferase domain and shows the residues required for catalytic activity. Residues from the other monomer involved in the catalytic mechanism are shown in purple.

The crystal structures for human, avian, *T. maritima* and *M. tuberculosis* (Greasley et al. 2001; Wolan et al. 2002; Wolan et al. 2003; Axelrod et al. 2008; Le Nours et al. 2011) show a high degree of structural similarity (Figure 3.5). However, prokaryote ATIC structures of *M. tuberculosis* and *Thermotoga maritima* show a 90° shift of the inosine monophosphate cyclohydrolase domain compared to eukaryote structures from human and avian ATIC (Axelrod et al. 2008). This rearrangement is believed to alter the interaction between the two domains and could result in a difference in coupling mechanisms between inosine monophosphate cyclohydrolase and the amino-4-imidazolecarboxamide ribonucleotide transferase domains (Axelrod et al. 2008).

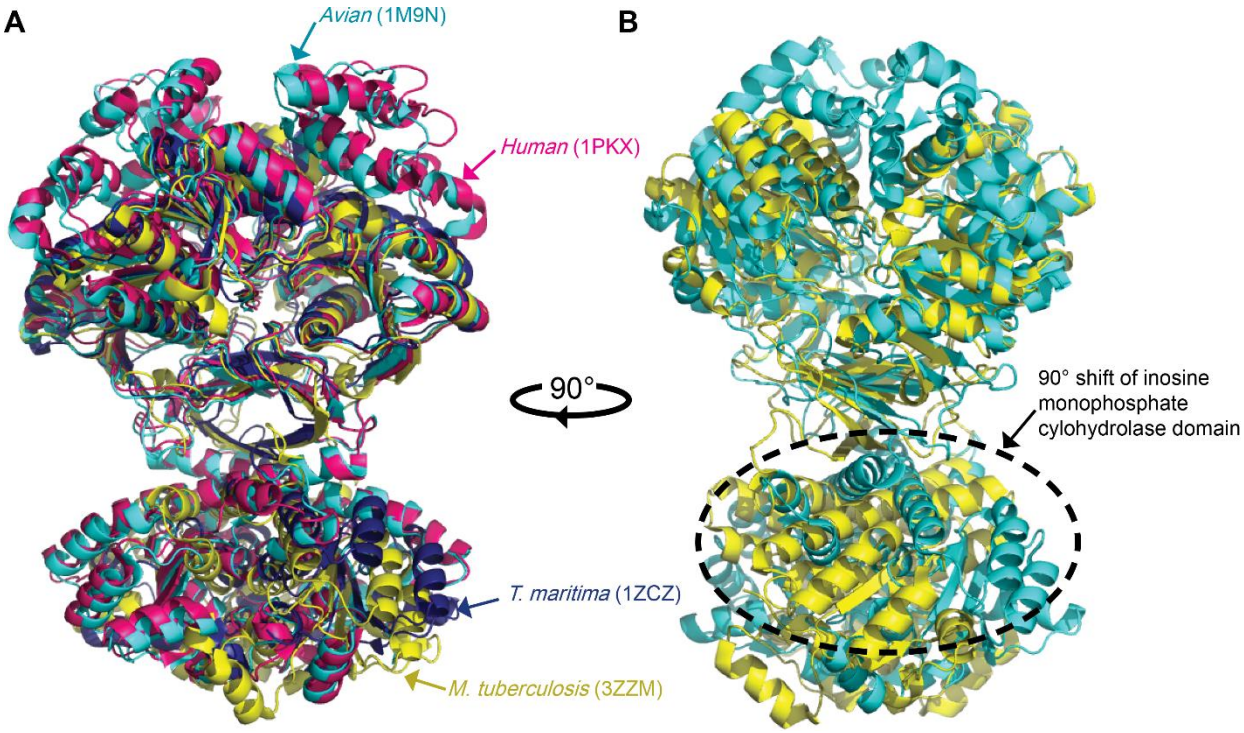



**Figure 3.5** Alignment of prokaryote and eukaryotic adenyltransferase/IMP cyclohydrolase structures. **A** shows the alignment of avian (PDB ID: 1M9N), human (PDB ID: 1PKX), *T. maritima* (PDB ID: 1ZCZ) and *M. tuberculosis* (PDB ID: 3ZZM). **B** is the alignment of only *the M. tuberculosis* and avian adenyltransferase/IMP cyclohydrolase to show the 90° shift in orientation of the inosine monophosphate cyclohydrolase domain.

### 3.1.5 Overview of the Chapter

There are few studies that report on *C*Lso enzyme function or structure thus, it is not known what function annotated proteins in the *C*Lso genome actually undertake. Furthermore, inconsistencies in genome wide annotations could cause inaccuracies in determining the function of an annotated protein. With Chapter Two demonstrating expression of the putative pathogenicity factor, adenyltransferase/IMP cyclohydrolase, in both psyllid and *in planta*, the question arose as to what role this enzyme plays in survival or pathogenicity of *C*Lso and whether it confers any adaptive advantage to the psyllid host. Thus, this Chapter characterises the adenyltransferase/IMP cyclohydrolase to determine whether the annotated protein exhibits the known structure and function comparable to other bacterial homologues. I describe successful expression and purification of soluble *C*Lso ATIC via the co-expression of *E. coli* chaperone proteins. The first solution structure for a protein from *C*Lso is presented with evidence the protein is stable and folded in solution. Although ATIC activity was not definitively proven, the results described here contribute significantly to characterizing *C*Lso proteins using the chaperone expression approach (described further in Chapter Four) and highlights ambiguities found in previously published studies on refolded proteins from Liberibacter species.

## 3.2 Results and Discussion

### 3.2.1 *C*Lso ATIC retains key residues required for activity

As endosymbiotic relationships have been known to undergo degenerative rather than adaptive genome evolution (van Ham et al. 2003), it was of interest to determine whether *C*Lso (a bacterial endosymbiont of *B. cockerelli* (psyllid)) ATIC retains key residues required for activity as changes in these residues could attenuate enzyme activity. A multiple sequence alignment was generated between ATIC of *C*Lso, *T. maritima, M. tuberculosis, Campylobacter jejuni, Staphylococcus. lugdenesis* and *Liberibacter crescens* (Figure 3.6) using CLUSTAL Omega. The alignment was used determine whether there are any significant changes in amino acid residues that could affect the functionality of the protein. The alignment showed that the two key catalytic residues Lys289 and His290 required for 5-amino-4-imidazolecarboxamide ribonucleotide transferase activity are conserved. Several other

conserved residues essential for AICAR binding and formyl group transfer are also conserved and include: Arg227, Tyr228, Gly229, Glu342, Ser431, Asn432, Arg452, Phe486, Asp491, Ser510 and Arg533. The residues Ser22, Lys26, Thr46, Thr49, Lys78, Thr79, Tyr117, Asp137, Ile138, Gly139, and Lys149 are required for inosine monophosphate cyclohydrolase activity and are also conserved. These residues are largely responsible for orientating the 4-carboxamide conformation from binding AICAR in the 5-amino-4-imidazolecarboxamide ribonucleotide transferase active site to that that promotes the intramolecular cyclization to form inosine monophosphate. Interestingly, a normally conserved region spanning from Ser483 to Phe489 which is required to provide a hydrophobic environment for the pterin ring of folate has a conservative mutation at position 487 from Phe to a non- polar Tyr in both *C*Lso and *L. crescens*. This Tyr487 could potentially play a role in destabilising hydrophobicity and overall AICARFT activity. As *C*Lso ATIC retains key residues for both 5-amino-4-imidazolecarboxamide ribonucleotide transferase and inosine monophosphate cyclohydrolase activity, it is hypothesised that the enzyme would retain the expected catalytic activity. 47

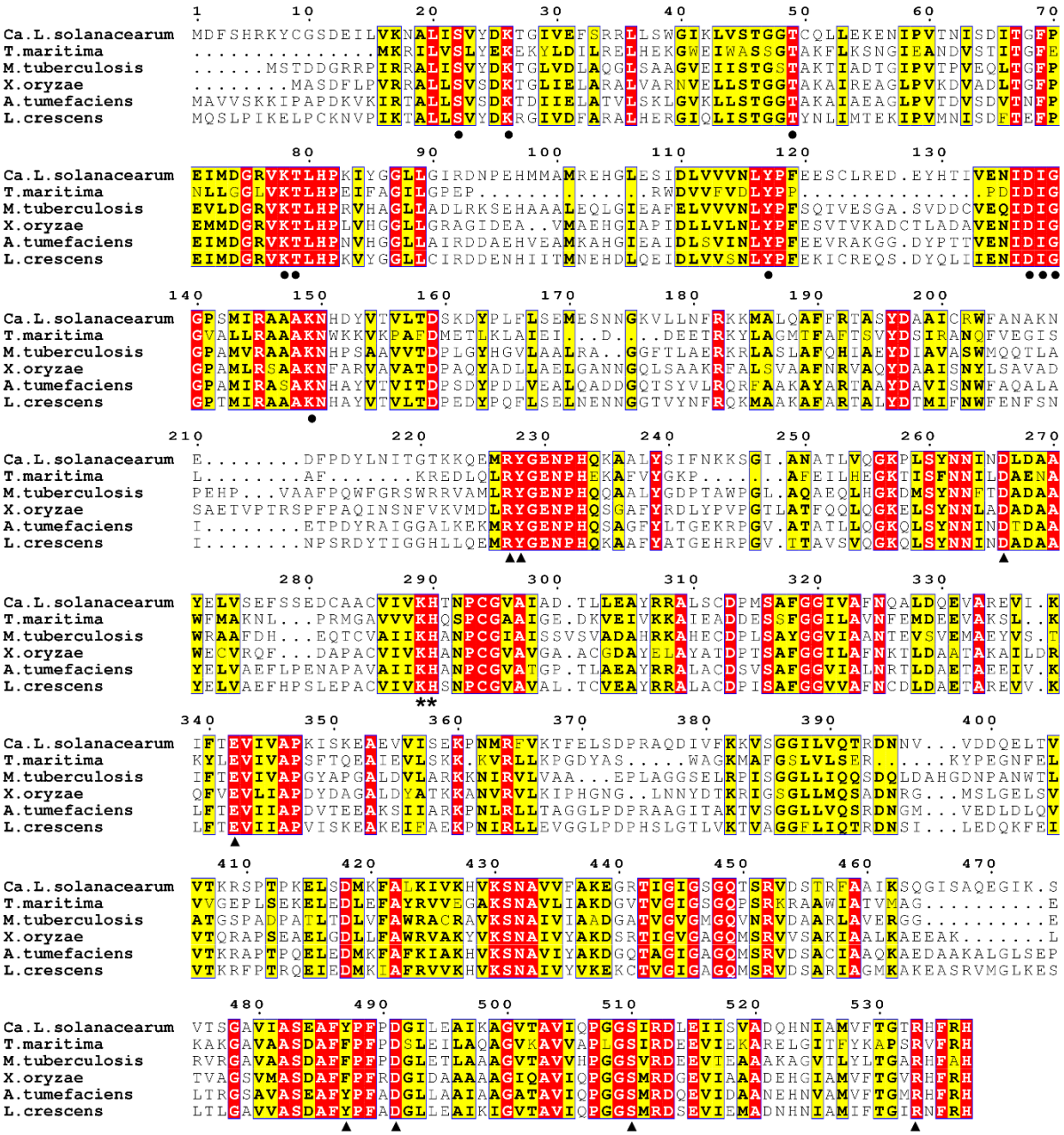



**Figure 3.6:** Multiple sequence alignment of *C*Lso ATIC with other bacterial homologues. The key catalytic residues responsible for 5-amino-4-imidazolecarboxamide ribonucleotide transferase activity are labelled with a black asterisk (*). Other conserved residues are indicated by ●. Black triangles (▲) indicate residues required for inosine monophosphate cyclohydrolase activity. The sequences used in the alignment were from *Thermotoga maritima* (AKE27158.1), *Ca*. L. solanacearum (KJZ81084.1), *Mycobacterium tuberculosis* (ALB18099.1), *Xanthomonas oryzae* (QBG84760.1), *Agrobacterium tumefaciens* (QAA98650.1) and *Liberibacter crescens* (WP_015273649.1).

### 3.2.2 Expression and purification of ATIC

*3.2.2.1 ATIC is insoluble when recombinantly expressed in E. coli*

In an attempt to produce soluble *C*Lso ATIC, the protein was initially overexpressed in *E. coli* BL21(DE3) cells with a N-terminal $His_6$-tag (Figure 3.7). Expression in an *E. coli* host system was utilised as *C*Lso is a Gram-negative alpha proteobacteria similar to *E. coli*, thus the system could provide similar molecular conditions to *C*Lso for protein. However, the expressed *C*Lso ATIC was insoluble (Figure 3.8A). SDS-PAGE analysis showed that *C*Lso ATIC was clearly overexpressed as an ~59-kDa insoluble protein, consistent with its predicted molecular mass of 60 kDa. Efforts to optimize expression conditions included varying the concentration of IPTG and growth temperature. Additionally, the plasmid was transformed into *E. coli* strains: BL21(DE3)pLysS, Origami, C41(DE3), C43(DE3), C41(DE3)pLysS and C43(DE3)pLysS and expression was induced with IPTG (Chapter Eight, Table 8.2 for expression strains and Section 8.6.1.1 for expression conditions). However, the *C*Lso ATIC protein was insoluble when expressed in these strains, with SDS-PAGE and western blot confirming that ATIC was not present in the soluble fraction (data not shown).

In order to determine whether the two domains were soluble when expressed as individual domains, *purH* was split to express each domain separately. The domains were overexpressed in *E. coli* BL21(DE3) cells.

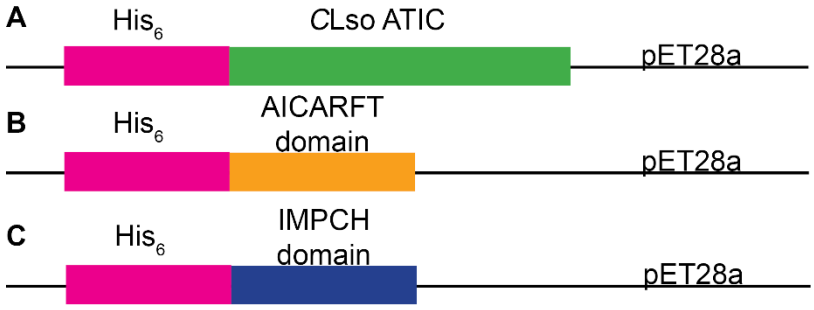



**Figure 3.7:** Schematic representation of the *C*Lso-ATIC construct. **A** is *C*Lso ATIC with a N-terminal $His_6$. **B** is the construct used for the amino-4-imidazolecarboxamide ribonucleotide transferase (AICARFT) domain expression with an N-terminal $His_6$ and **C** is the construct for the inosine monophosphate cyclohydrolase (IMPCH) domain with an N-terminal $His_6$.

Both the amino-4-imidazolecarboxamide ribonucleotide transferase domain and the inosine monophosphate cyclohydrolase domain were insoluble when expressed separately (Figure 3.8B). SDS-PAGE analysis showed that the *C*Lso ATIC inosine monophosphate cyclohydrolase domain was overexpressed as an ~30 kDa insoluble protein whilst the *C*Lso ATIC 5-amino-4-imidazolecarboxamide ribonucleotide transferase domain truncation was overexpressed as a ~32 kDa insoluble protein (Figure 3.8B).

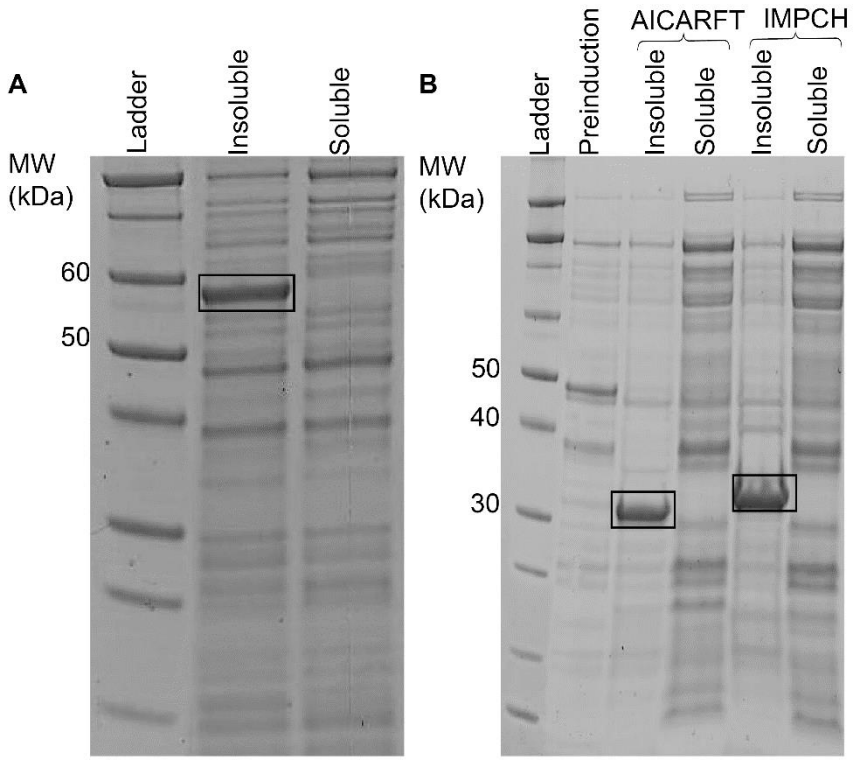



**Figure 3.8:** SDS-PAGE analysis of *C*Lso ATIC. **A** shows the expression of *C*Lso ATIC present in as an insoluble protein whilst **B** shows the expression of the individual ATIC domains: 5-amino-4-imidazolecarboxamide ribonucleotide transferase (AICARFT) domain and the inosine monophosphate cyclohydrolase (IMPCH) domain also showing that the separate domains are insoluble.

### **3.2.2.2** *ATIC with fusion tags*

In another attempt to express *C*Lso ATIC as a soluble protein, the gene encoding ATIC from *C*Lso was cloned into a suite of ten pOPIN expression vectors (Chapter Eight, Section 8.3.2 and Section 8.3.2.1). These constructs are detailed in Chapter Eight (Section 8.3.2.1), briefly, fusion tags are widely known to aid in improving the solubility of proteins (Costa et al. 2014; Malik 2016). Here, a suite of ten pOPIN expression vectors were used which expressed a $His_6$ tag followed by the fusion tag (for example: green fluorescent protein, maltose binding

protein, trigger factor), a 3C cleavage site for proteinase cleavage of the fusion and $His_6$ tag followed by the protein of interest, *C*Lso ATIC (Figure 3.9).


**Figure 3.9:** Schematic representation of the pOPIN fusion protein expression system.

The constructs were transformed into *E. coli* Tuner bacterial cells and cultured in lysogeny medium. The proteins were expressed and the cells were harvested and lysed. Soluble $His_6$TF-*C*LsoATIC was expressed as a ~110 kDa protein, consistent with its predicted molecular mass of 109 kDa, was only soluble when expressed with the trigger factor fusion tag (pOPIN-TF plasmid).The 109 kDa protein was purified by a combination of immobilized metal affinity chromatography and size exclusion chromatography. The increasing purity of the $His_6$TF-*C*LsoATIC protein after each purification cycle is shown in Figure 3.8A.

Unfortunately, size exclusion chromatography (Figure 3.10B) and further biophysical characterisation by analytical ultracentrifugation determined that purified $His_6$TF-*C*LsoATIC protein was present in solution as high molecular weight oligomer of $His_6$TF-*C*LsoATIC (Figure 3.10C and D) as it eluted in the void volume. Sedimentation velocity analysis by the

continuous sedimentation distribution method show a distribution range up to 150 S (Figure 3.10D).

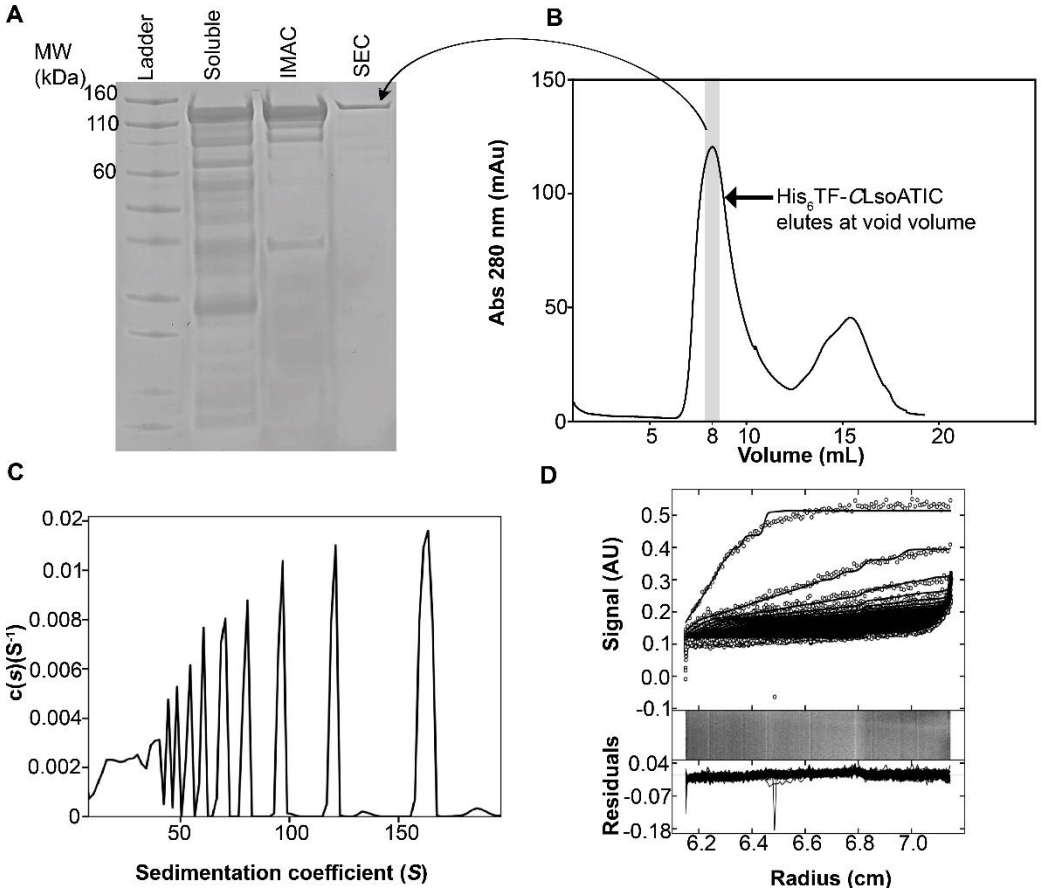



**Figure 3.10:** Purification and sedimentation velocity analysis of $His_6$TF-*C*LsoATIC. **A** shows the increasing purity of $His_6$TF- *C*LsoATIC during the purification process whilst **B** shows the elution of the protein at the void volume of the size exclusion chromatography column. **C** shows the continuous sedimentation plotted as a function of sedimentation coefficient (Svedberg, S). The fit was obtained using a resolution of 100 species between $S_{min}$ of 1S and $S_{max}$ of 200S with P = 0.95 vbar = 0.73. The r.m.s.d and runs test average for the fit were calculated to be 0.0047 and 1.91 respectively. D is the absorbance at 280 nm plotted as a function of radial position from the axis of rotation (cm). The raw data are represented as open symbols (o) plotted over a time interval, every third scan is shown. Raw data are overlaid with the nonlinear least squares best fit to a continuous distribution model. Residuals for the nonlinear least squares best fit are also shown. The continuous sedimentation plotted as a function of sedimentation coefficient (Svedberg, S). The fit was obtained using a resolution of 100 species between $S_{min}$ of 1S and $S_{max}$ of 200S with P = 0.95 vbar = 0.73 (Figure 3.10C). The r.m.s.d and runs test average for the fit were calculated to be 0.0047 and 1.91 respectively. The raw data are represented as open symbols (o) plotted over a time interval. Raw data are overlaid with the nonlinear least squares best fit to a continuous distribution model.

This oligomer form of $His_6$TF-*C*LsoATIC was observed even after cleavage of the trigger factor fusion tag and after denaturation with 8 M urea and 6 M guanidine hydrochloride and refolding of the protein (data not shown). This oligomeric *C*Lso ATIC in solution highlights a major limitation to current *Ca*. Liberibacter research in that most purification strategies reported to date require the denaturation and refolding of the *Ca*. Liberibacter recombinant protein with very little characterization to ensure the integrity of the protein (Liu et al. 2017; Ravindran et al. 2018). Thus, it is not possible to determine whether the protein is truly non-functional since the protein could be persisting in solution in a non-native state preventing the binding of substrates due to incorrect refolding of the protein.

Various approaches were investigated to prevent aggregation in the soluble phase, including purification at pH 5 – 9.5, various salt concentrations (0 – 1.5 M) and with combinations of arginine and proline which are known to prevent protein aggregation (Das et al. 2007; Lang and Rudolph 2009) as well as the detergents 0.1% Triton-X and n-dodecyl beta-d-maltoside (DDM) (see Chapter Eight, Section 8.6.1.3). All buffers used contained 0.1 mM of the reducing agent tris(2-carboxyethyl)phosphine. *E. coli* Tuner(DE)pLacI cells used for the expression of $His_6$TF-*C*LsoATIC were supplemented with addtives in the growth media, which included sorbitol, arginine, sodium chloride, glycerol, arginine, benzyl alcohol and ethylene glycol to provide chemical chaperones in order to prevent $His_6$TF-*C*LsoATIC aggregation (Prasad et al. 2011) (Chapter Eight, Section 8.6.1.3).

Finally, the protein was purified under denaturing conditions of 8 M urea and 6 M guanidine hydrochloride. The denaturant was added to the resuspension buffer and added to the immobilised metal affinity purification buffers, after purification, the protein was dialysed against the size exclusion buffer which did not contain denaturant in order to refold the protein. Size exclusion chromatography determined that the refolded $His_6$TF-*C*LsoATIC eluted in the void volume of the column consistent with the elution of aggregated protein or very large protein complexes (> 1, 000, 000 Da).

#### *3.2.2.3 ATIC co-expression with chaperones*

Nachappa et al. (2012) conducted transcriptomic analysis of *C*Lso infected psyllids and found that the genes for the GroEL, DnaK and GrpE chaperone proteins were highly expressed in

*C*Lso. I reasoned that these chaperones could potentially aid in *C*Lso protein folding *in vitro*. Therefore, co-expression of *C*Lso ATIC with *E. coli* chaperones was investigated in order to determine whether chaperone co-expression could mimic *in vivo* *C*Lso protein folding conditions in order to obtain soluble and functional protein. The *C*Lso ATIC protein with a C-terminal $His_6$ was overexpressed in *E. coli* BL21(DE3) that also contained one of a range of additional plasmids encoding *E. coli* chaperone proteins (Figure 3.11) (Table 3.1) (see Chapter Eight for chaperone expression method).


**Figure 3.11:** Schematic representation of the *C*Lso ATIC protein for chaperone co-expression.

**Table 3.1:** Plasmids used for co-expression of chaperones with *C*Lso proteins.

| Plasmid | Chaperone | Expression Host | Chaperone molecular weight (kDa) | Promoter | Inducer |
|---|---|---|---|---|---|
| pGRO7 | GroES<br>GroEL | *E.coli* BL21(DE3) | 10<br>60 | *araB* | L-arabinose |
| pKJE7 | dnaK<br>dnaJ<br>grpE | *E.coli* BL21(DE3) | 70<br>40<br>22 | *araB* | L-arabinose |
| pTf16 | Trigger factor | *E.coli* BL21(DE3) | 56 | *araB* | L-arabinose |

*C*Lso ATIC was successfully expressed as a soluble 59 kDa protein when co-expressed with the DnaK/DnaJ/GrpE chaperone proteins, consistent with the predicted molecular mass of 60 kDa (Figure 3.12A). *C*Lso ATIC was then purified to homogeneity using a combination of anion exchange (Q Sepharose and Q Sepharose high performance strong quaternary ammonium resin) and size exclusion chromatography (Chapter Eight, Section 8.7.2) (Figure 3.12B).

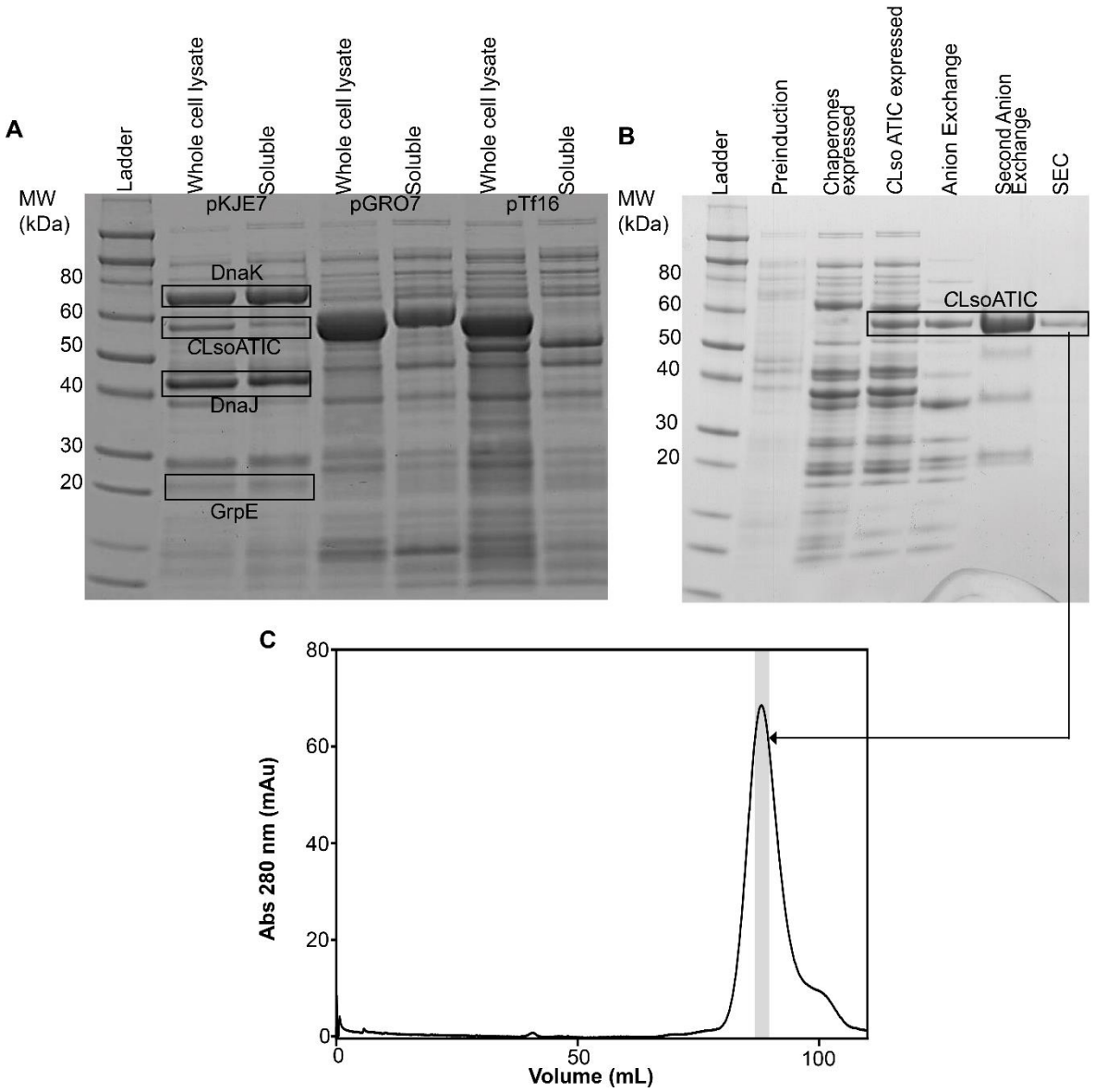



**Figure 3.12:** Chaperone co-expression of *C*Lso ATIC. **A** shows the expression of soluble *C*Lso ATIC in the pKJE7 plasmid with the DnaK, DnaJ and GrpE chaperones. **B** shows the purification of the protein through a combination of anion exchange and size exclusion chromatography whilst **C** is the size exclusion chromatography showing *C*Lso ATIC elutes at approximately 85 mL.

### 3.2.3 Activity Assay

The kinetic assay as described by Mueller and Benkovic (1981) and Black et al. (1978) was trialled however after a number of attempts (Chapter Eight, Section 8.9.1), there was no measurable rate of reaction for either catalytic site. Briefly, the assay measures the forward 5-aminoimidazole-4-ribonucleotide transferase activity *via* the addition of AICAR, 10-

formyltetrahydrofolate, potassium chloride and *C*Lso ATIC, and measuring the formation of tetrahydrofolate at 298 nm. The inosine monophosphate cyclohydrolase reaction was monitored by the synthesis of inosine monophosphate at 248 nm after the addition of *C*Lso ATIC to 5-formylaminoimidazole-4-carboxamide ribonucleotide and tetrahydrofolate (Mueller and Benkovic 1981).

*C*Lso ATIC was found to be non-functional, albeit a positive control could not be sourced for the assay. Structural insights indicate that although the functionality of *C*Lso ATIC could be dependent on protein dimerisation, the catalytic activity of the AICARFT and IMPCH domains have been shown to retain some activity in the monomer form (Rayl et al. 1996). Further, lack of activity could be the result of the environment such as temperature or pH. Thus, *C*Lso ATIC could be functional at a contrasting pH or temperature compared to those described for human ATIC (Mueller and Benkovic 1981). Alternatively, the lack of observable ATIC activity could potentially be a result of gene mis-annotation. Further structural characterization of *C*Lso ATIC would be required to determine whether a mis-annotation of the gene has occurred during automated annotation pipelines. Another hypothesis is that cellular conditions contrast to those used *in vitro*, it is possible that high concentrations of ATIC are present within the *C*Lso cell and could allow for dimerisation, thus, the low concentration of *C*Lso ATIC used within the assay (0.3 mg/mL) may not reflect the cellular conditions. Additionally, *C*Lso ATIC could have lost the ability to synthesise and metabolise purines (Duan et al. 2009; Lin et al. 2011).

*C*Lso ATIC has been successfully expressed and purified, a major advancement for the ability to investigate future proteins from the organism. However, although it is unclear, the enzyme did not produce any catalytic activity for either the 5-aminoimidazole-4-ribonucleotide transferase or inosine monophosphate cyclohydrolase domains. With dimerization being required for enzyme activity, it is necessary to understand whether the lack of activity might be due to a change in structure of the protein, thus preventing the correct formation of the two catalytic sites.

### 3.2.4 Solution structure of ATIC

*3.2.4.1 Secondary structure as determined by circular dichroism spectroscopy*

Next, I tested whether the protein is folded. Certain structural elements of proteins, such as α-helices and β-sheets, have characteristic far UV CD spectra (180 – 240 nm) (Greenfield, 2006). For instance, α-helices produce minima at 222 nm and 208 nm and a maximum at 193 nm, while β-sheets produce a minimum at 218 nm and a maximum at 195 nm. Disordered proteins, on the other hand, show very little ellipticity above 210 nm and a minimum at 195 nm, making far-UV CD spectroscopy a useful method for determining whether a protein is folded or not. For *C*Lso ATIC, the spectrum recorded shows that the enzyme is folded (Figure 3.13). A broad minimia is apparent between 208 nm and 222 nm, which is characteristic of enzymes composed of both alpha helices and beta strands (Figure 3.13).

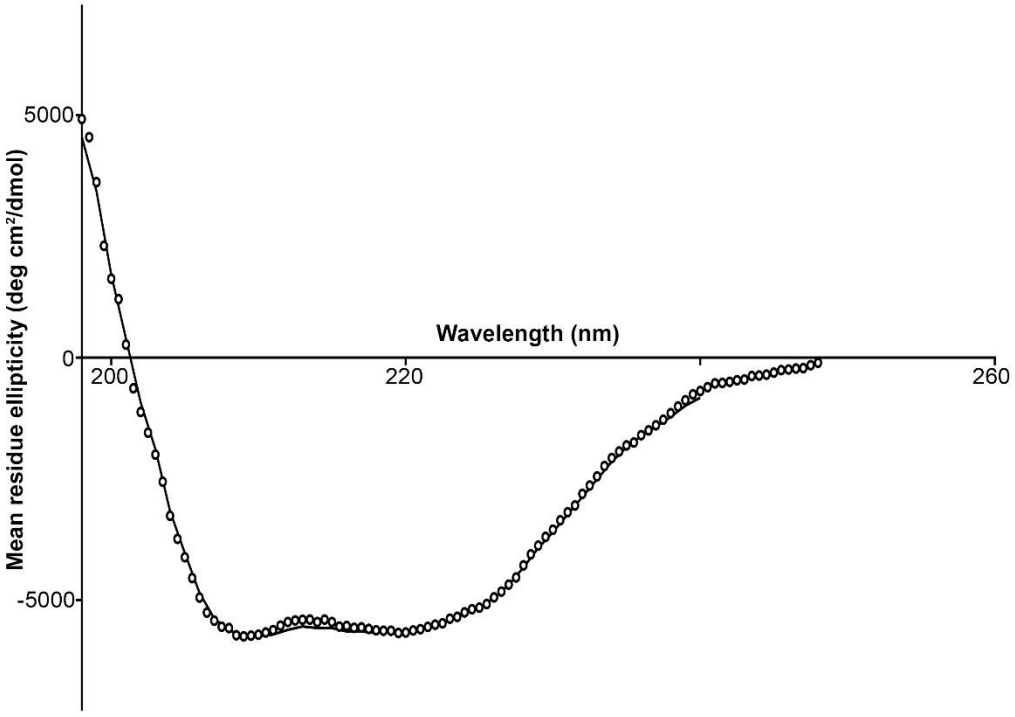



**Figure 3.13:** Far-UV CD spectra of *C*Lso ATIC. *C*Lso ATIC was used at 0.05 mg/ml in water in a 1 mm cuvette. The spectra were collected between 180 and 240 nm. The ellipticity data were converted to the mean residue ellipticity. Data (circles) are overlaid with a fit (line) against the SMP180 reference set of proteins using the CONTIN algorithm within DichroWeb (Whitmore and Wallace 2008).

*C*Lso ATIC spectra was then compared to the SMP180 reference set of proteins using the CONTIN algorithm (Provencher and Gloeckner 1981; van Stokkum et al. 1990) within DichroWeb (Whitmore and Wallace 2008) to estimate the secondary structure composition of the protein. The purified *C*Lso ATIC was composed of 30% alpha helices, 15% beta strands, 25% turns and 30% unordered structure by CD analysis. These values are similar to the structure of the protein predicted by PSI-BLAST based secondary structure PREDiction (PSIPRED) (35% alpha helices, 17% beta strands and 48% random coils. CD results correlated with the predicted secondary structure of *C*Lso ATIC, thus, it was determined that the protein was folded in solution.

To date, there are no studies that have provided a means of overcoming the difficulties in *C*Lso protein expression and purification. Evidence of folded *C*Lso ATIC via the use of chaperone co-expression is a crucial for further investigation of the enzyme function and structure.

*3.2.4.2 Quaternary structure as determined by analytical ultracentrifugation*

Circular dichroism determined *C*Lso ATIC was folded in solution, however, the ability to purifiy soluble folded protein from the organism has presented a number of opportunities to investigate and validate the proteins annotated in the genome and to determine whether the enzymes function as predicted. Therefore, analytical ultracentrifugation was used in order to investigate the oligomeric state of the protein in order to see whether the structure in solution is similar to other adenyltransferase/IMP cyclohydrolase homologues.

Excitingly, *C*Lso ATIC expressed with chaperones exists as a single monomeric species in solution. Although a monomer at the particular concentration tested, an increase in protein concentration can result in the formation of higher order oligomeric structures (Merten et al. 2011), thus *C*Lso ATIC concentration varied from 0.2 mg/mL up to 0.6 mg/mL in order to determine whether ATIC dimerization occurred at higher concentrations. Increasing concentrations of *C*Lso ATIC did not aid in dimerisation (Figure 3.14). Data were fitted to a continuous sedimentation and mass coefficient [c(*s*) and c(*M*)] model, resulting in a distinct single, narrow and symmetrical peak with a sedimentation coefficient of 4 S and a mass of 56 kDa which is close to the theoretical monomeric mass of 59,332 Da, compared to the

theoretical dimeric mass of 118,664 Da. Interestingly, ATIC from other species exists in a monomer-dimer equilibrium, with the dimer required for AICARFT transferase activity (Vergis et al. 2011). This raises the question as to whether *C*Lso ATIC is functional as a monomer or as a result of genetic drift, the enzyme is no longer functional in *C*Lso.

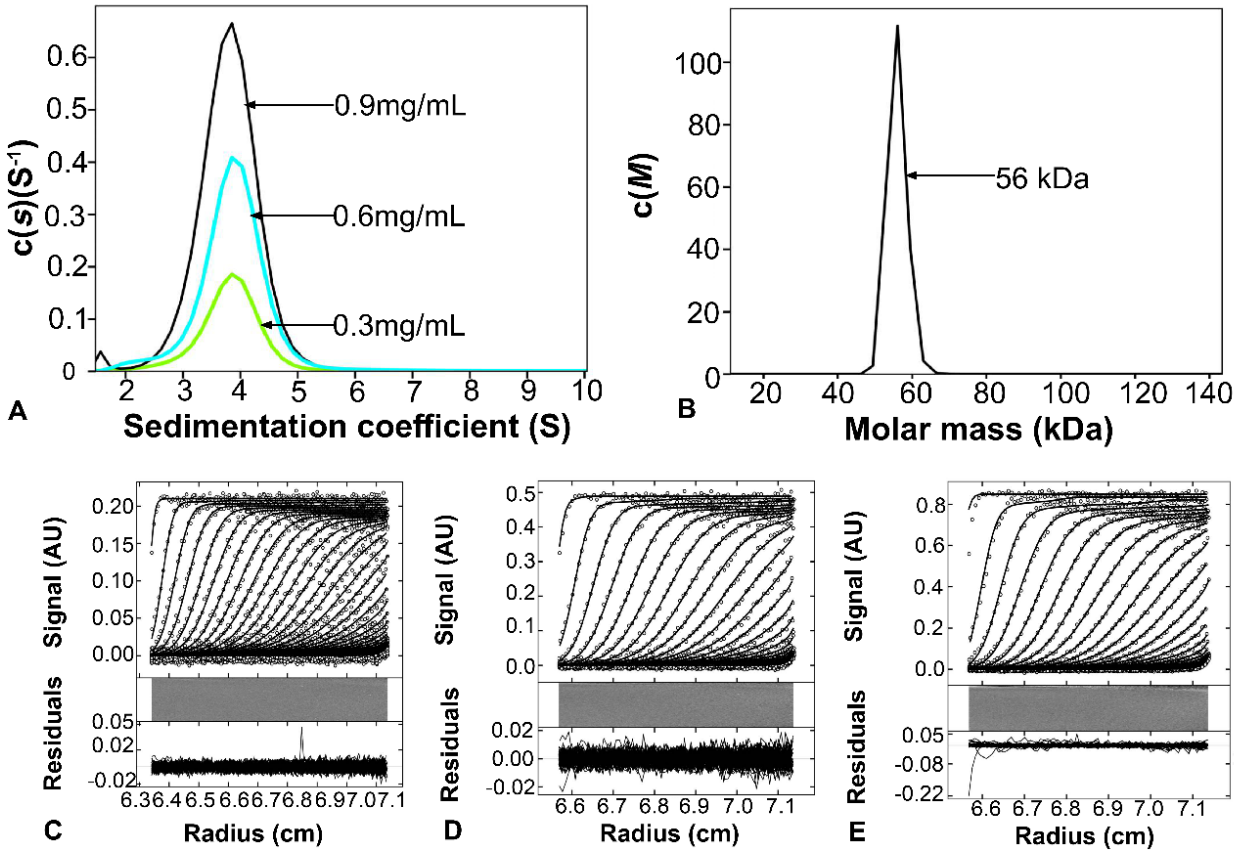



**Figure 3.14:** Sedimentation velocity analysis of *C*Lso ATIC co-expressed with chaperones. **A** is the continuous sedimentation plotted as a function of sedimentation coefficient (Svedberg, S). The fit was obtained using a resolution of 100 species between $S_{min}$ of 1S and $S_{max}$ of 10S with P = 0.95 vbar = 0.73. The r.m.s.d. and runs test average for the fit were calculated to be 0.0047 and 1.91, respectively. **B** is the molecular weight distribution of ATIC plotted as a function of molar mass (kDa). The fit was obtained using a resolution of 100 species between 0 – 140 kDa with P = 0.95 vbar = 0.73. The r.m.s.d. and runs test average for the fit were 0.0062 and 1.2, respectively. **C.D.E**. is the absorbance at 280 nm plotted as a function of radial position from the axis of rotation (cm). The raw data are represented as open symbols (o) plotted over a time interval. Raw data are overlaid with the nonlinear least squares best fit to a continuous distribution model. Residuals for the nonlinear least squares best fit are also shown. **C** are the residuals for at 0.3 mg/mL. **D** are the residuals at 0.6 mg/mL, **E**. are the residuals at 0.9 mg/mL. Every third scan is shown.

*3.2.4.3 Small angle X-ray Scattering (SAXS) analysis*

Analytical ultracentrifugation determined that *C*Lso ATIC was a monomer in solution, which contrasts to other known structures of the enzyme which form a dimer in solution. To investigate the shape of the enzyme in solution and to further support the data for *C*Lso ATIC being a monomer in solution SAXS data was collected for the enzyme. At small values of q (the scattering vector), the linear portion of the Guinier plot (ln(I) vs $q^2$) can be used to extrapolate back to I(0) (scatter intensity at the origin) and to estimate $R_g$ (the radius of gyration) (Jullien 1992). The linearity of the Guinier plots suggests that the samples were free from aggregation or inter-particle effects (Figure 3.15A inset). Scattering data are presented as an intensity plot (Figure 3.15A).

An indirect Fourier transform, performed using GNOM, yields the function *P(r)*, which gives the relative probabilities of distances between scattering centres and the maximum dimension of the particle, $D_{max}$ (Svergun 1992). The maximum dimension of the particle was ~71 Å (Figure 3.15B). The molecular mass calculated for *C*Lso ATIC extracted from the SAXS data is 35 kDa which differs from the apparent molecular mass of 56 kDa derived from sedimentation velocity analysis, however, the mass is more consistent with a monomer than a dimer. The *P(r)* distribution shows a single peak that is symmetrical in shape, with no positively skewed tail at long distances (Figure 3. 15B). The Kratky plot shows a bell shaped plot which further indicates the protein is folded with some flexibility in solution as the Kratky plot does not return to the baseline (Figure 3.15C). This indicates that *C*Lso ATIC is globular in shape and not elongated in solution. The structural parameters are presented in Table 3.2.

In an effort to determine whether the structure in solution correlates with known ATIC structures*, C*Lso ATIC experimental scattering data was compared to the theoretical scattering generated from the atomic coordinates of the *M. tuberculosis* adenyltransferase/IMP cyclohydrolase (PDB: 3ZZM) using CRYSOL (Svergun et al. 1995). The experimental scattering was compared to the dimeric and monomeric *M. tuberculosis* ATIC (Figure 3.15A). Although both ATIC oligomeric states provided poor fits to the theoretical scattering data, with chi squared values of 40.13 and 4.41 for the dimer and monomer respectively. SAXS REFinement through FLEXibility (SREFLEX) was used to determine whether varying conformations of the

monomer could improve the fit, which resulted in an improved fit to the data of $\chi^2 = 0.44$ . Similarly, the $D_{max}$ for both the ATIC monomer ($D_{max}$ = 100 Å) and dimer ($D_{max}$ = 102 Å) did not correlate with the experimental $D_{max}$ of 71 Å, whilst the confomer had a consistent $D_{max}$ of 67 Å.

*C*Lso ATIC appears to have different conformation in solution to the known structure of *M. tuberculosis* ATIC. The experimental $D_{max}$ indicates that *C*Lso ATIC may have a more compacted structure, which in turn could prevent dimerization of the protein, and thus be the reason for *C*Lso ATIC existing as a monomer in solution.

**Table: 3.4:** Data collection and analysis statistics from SAXS

| **SAXS data analysis** | **Adenyltransferase/IMP cyclohydrolase** |
|---|---|
| *I*(0) ($cm^{-1}$) (Guinier analysis) | 0.012 ±4.2e-05 |
| $R_g$ (Å) (Guinier analysis) | 2 ± 1.341 |
| *I*(0) ($cm^{-1}$) (*P*(*r*) analysis) | 0.012 |
| $R_g$ (Å) (*P*(*r*) analysis) | 22.01 |
| $D_{max}$ (Å) | 70.6 |
| Porod volume ($Å^{-3}$) | 50500 |
| Molar mass (Porod volume, kDa) | 30300 |
| Molar mass (SAXSMoW2*, kDa) | 34250 |
| Calculated molar mass from sequence (kDa) | 56 000 |
| **SAXS data collection parameters** | |
| Instrument | Australian Synchrotron SAXS/WAXS beamline |
| detector | PILATUS 1M (Dectris) |
| wavelength (Å) | 1.0332 |
| Maximum flux at sample | 8 x $10^{12}$ photons per second at 12 keV |
| Camera length (mm) | 1600 |
| Q range ($Å^{-1}$) | 0.006-0.5 |
| Exposure time | Continuous 1 second frame measurements |
| Sample configuration | SEC-SAXS with co-flow |
| Sample temperature (°C) | 12 |

*http://saxs.ifsc.usp.br/

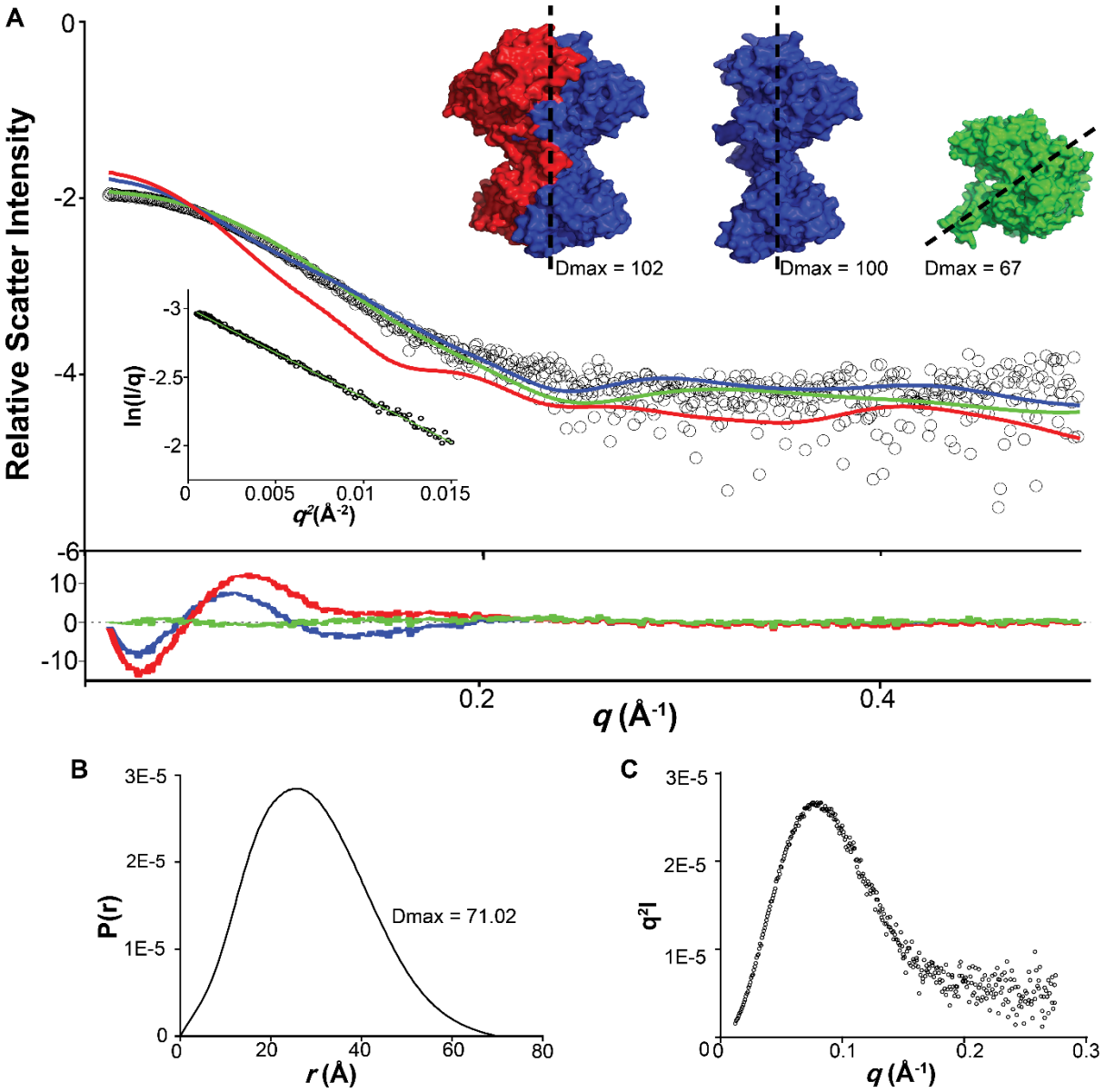



**Figure 3.15:** SAXS data for *C*Lso ATIC. **A** is the comparison of the experimental scattering and theoretical scattering data. The experimental scattering profile of *C*Lso ATIC (presented as open circles, o) overlaid with theoretical scattering profile of the *M. tuberculosis* dimeric structure in red (inset) ($\chi^2$ = 40.13) and the monomer in blue (inset) ($\chi^2$ = 4.41) whilst the theoretical scattering profile of the SREFLEX conformer structure is in green (inset with a $\chi^2$ = 0.44). The Guinier plot shown as an insert, highlights linearity at low q which provides confidence that the data is free of aggregation or inter-particle interference. **B** is the pairwise distribution (*P(r)*) plot displaying the maximum inter-particle dimension ($D_{max}$) as 71 Å. **C** is the Kratky plot shows a bell shaped curve, which indicates the protein is folded.

## 3.3 Summary

For the first time, a protein from *C*Lso has been purified without the use of denaturation and refolding. Size exclusion chromatography and analytical ultracentrifugation highlighted the presence of soluble aggregated $His_6$*C*LsoATIC-TF after refolding in both urea and guanidine hydrochloride which demonstrated the limitations of previous studies and the need for extensive characterisation of *C*Lso proteins before functional assays can be undertaken.

The use of chaperone co-expression resulted in soluble, purified ATIC. ATIC was folded and existed as a monomer in solution. Interestingly, SAXS analysis indicated that the protein could have an altered conformation when compared to known ATIC monomer or dimer states. This altered conformation was hypothesized to potentially prevent dimerization which in turn could account for the lack of observable activity. This is the first attempt to characterise proteins from *C*Lso and provided a potential approach for obtaining soluble functional protein from the organism which is investigated further in Chapter Four.

## 3.4 References

Axelrod, H.L., McMullan, D., Krishna, S.S., Miller, M.D., Elsliger, M.A., Abdubek, P., Ambing, E., Astakhova, T., Carlton, D., Chiu, H.J., and Clayton T. 2008. Crystal structure of AICAR transformylase IMP cyclohydrolase (TM1249) from *Thermotoga maritima* at 1.88 Å resolution. Proteins: Structure, Function, and Bioinformatics 71(2):1042-9.

Beardsley, G.P., Rayl, E.A., Gunn, K., Moroson, B.A., Seow, H., Anderson, K.S., Vergis, J., Fleming, K., Worland, S., Condon, B., Davies, J. 1998. Structure and functional relationships in human *purH*. Purine and Pyrimidine Metabolism in Man IX 1998 (pp. 221-226). Springer, Boston, MA.

Becerra, A., and Lazcano, A. 1998. The role of gene duplication in the evolution of purine nucleotide salvage pathways. Origins of Life and Evolution of the Biosphere 28(4-6):539-53.

Bulock, K.G., Beardsley, G.P., and Anderson, K.S. 2002. The kinetic mechanism of the human bifunctional enzyme ATIC (5-amino-4-imidazolecarboxamide ribonucleotide transformylase/inosine 5'-monophosphate cyclohydrolase) a surprising lack of substrate channeling. Journal of Biological Chemistry 277(25):22168-74.

Chatterjee, S., and Sonti, R.V. 2005. Virulence deficiency caused by a transposon insertion in the *purH* gene of *Xanthomonas oryzae* pv. *oryzae*. Canadian Journal of Microbiology 51(7):575-81.

Cheong, C.G., Wolan, D.W., Greasley, S.E., Horton, P.A., Beardsley, G.P., and Wilson, I.A. 2004. Crystal structures of human bifunctional enzyme ATIC in complex with potent sulfonyl-containing antifolates. Journal of Biological Chemistry. Journal of Biological Chemistry 13(2)1770.

Das, U., Hariprasad, G., Ethayathulla, A.S., Manral, P., Das, T.K., Pasha, S., Mann, A., Ganguli, M., Verma, A.K., Bhat, R., and Chandrayan, S.K. 2007. Inhibition of protein aggregation: supramolecular assemblies of arginine hold the key. PloS One 2(11):e1176.

Duan, Y., Zhou, L., Hall, D.G., Li, W., Doddapaneni, H., Lin, H., Liu, L., Vahling, C.M., Gabriel, D.W., Williams, K.P., and Dickerman, A. 2009. Complete genome sequence of citrus huanglongbing bacterium, '*Candidatus* Liberibacter asiaticus' obtained through metagenomics. Molecular Plant-Microbe Interactions 22(8):1011-20.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O., and Venter, J.C. 2006. Essential genes of a minimal bacterium. Proceedings of the National Academy of Sciences 103(2):425-30.

Greasley, S.E., Horton, P., Ramcharan, J., Beardsley, G.P., Benkovic, S.J., and Wilson, I.A. 2001. Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis. Nature Structural & Molecular Biology 8(5):402.

Greenfield, N.J. 2006. Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols 1(6):2876.

Hedstrom, L. 2009. IMP dehydrogenase: structure, mechanism, and inhibition. Chemical Reviews 109(7):2903-28.

Jenkins, A., Cote, C., Twenhafel, N., Merkel, T., Bozue, J., and Welkos, S. 2011. Role of purine biosynthesis in *Bacillus anthracis* pathogenesis and virulence. Infection and Immunity 79(1):153-66.

Kim, J.K., Am Jang, H., Won, Y.J., Kikuchi, Y., Han, S.H., Kim, C.H., Nikoh, N., Fukatsu, T., and Lee, B.L. 2014. Purine biosynthesis-deficient *Burkholderia* mutants are incapable of symbiotic accommodation in the stinkbug. The ISME journal 8(3):552.

Lange, C., and Rudolph R. 2009. Suppression of protein aggregation by L-arginine. Current Pharmaceutical Biotechnology 10(4):408-14.

Laue, T.M., and Stafford, W.F. 1992. Modern applications of analytical ultracentrifugation. Annual Review of Biophysics and Biomolecular Structure 28(1):75-100.

Le Nours, J., Bulloch, E.M., Zhang, Z., Greenwood, D.R., Middleditch, M.J., Dickson, J.M., and Baker, E.N. 2011. Structural analyses of a purine biosynthetic enzyme from *Mycobacterium tuberculosis* reveal a novel bound nucleotide. Journal of Biological Chemistry 286(47):40706-16.

Lebowitz J, Lewis MS, Schuck P. Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Science 11(9):2067-79.

Liechti, G., and Goldberg, J.B. 2012. Helicobacter pylori relies primarily on the purine salvage pathway for purine nucleotide biosynthesis. Journal of Bacteriology 194(4):839-54.

Lin, H., Han, C.S., Liu, B., Lou, B., Bai, X., Deng, C., Civerolo, E.L., and Gupta, G. 2013. Complete genome sequence of a Chinese strain of "*Candidatus* Liberibacter asiaticus". Genome Announcements 1(2):e00184-13.

Liu, H., Atta, S., and Hartung, J.S. 2017. Characterization and purification of proteins suitable for the production of antibodies against '*Ca*. L. asiaticus.' Protein Expression and Purification 139: 36–42.

Lobley, A., Whitmore, L., and Wallace, B.A. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. Bioinformatics 18(1):211-2.

McFarland, W.C., and Stocker, B.A. 1987. Effect of different purine auxotrophic mutations on mouse-virulence of a Vi-positive strain of *Salmonella dublin* and of two strains of *Salmonella typhimurium*. Microbial Pathogenesis 3(2):129-41.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. 1998. Large-Scale Identification of Virulence Genes from *Streptococcus pneumoniae*. Infection and Immunity 66(12):5620-9.

Prasad, S., Khadatare, P.B., and Roy, I. 2011. Effect of chemical chaperones in improving the solubility of recombinant proteins in Escherichia coli. Applied Environmental Microbiology 77(13):4603-9.

Provencher, S.W., and Gloeckner, J. 1981. Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20(1):33-7.

Ramsey, J.S., MacDonald, S.J., Jander, G., Nakabachi, A., Thomas, G.H., and Douglas, A.E. 2010. Genomic evidence for complementary purine metabolism in the pea aphid, *Acyrthosiphon pisum*, and its symbiotic bacterium *Buchnera aphidicola*. Insect Molecular Biology 19:241-8.

Ravindran, A., Saenkham, P., Gad-Levy, J., Tamborindeguy, C., Lin, H., Gross, D., and Pierson, E.A. 2018. Characterization of the serralysin-like gene of *Candidatus* Liberibacter solanacearum associated with potato zebra chip disease. Phytopathology 108(3): 327–335.

Rayl, E.A., Moroson, B.A., and Beardsley, G.P. 1996. The human *purH* gene product, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase cloning, sequencing, expression, purification, kinetic analysis, and domain mapping. Journal of Biological Chemistry 271(4):2225-33.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Molecular Microbiology 48(1):77-84.

Svergun, D.I., Barberato, C., and Koch, M.H. 1995. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. Journal of Applied Crystallography 28(6):768-73.

Toft, C., and Andersson, S.G. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. Nature Reviews Genetics 11(7):465.

Van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., Van Grondelle, R., and Groen, F.C. 1990. Estimation of protein secondary structure and error analysis from circular dichroism spectra. Analytical Biochemistry 191(1):110-8.

Vergis, J.M., and Beardsley GP. Catalytic mechanism of the cyclohydrolase activity of human aminoimidazole carboxamide ribonucleotide formyltransferase/inosine monophosphate cyclohydrolase. Biochemistry 43(5):1184-92.

Vergis, J.M., Bulock, K.G., Fleming, K.G., and Beardsley, G.P. 2001. Human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase/inosine 5'-monophosphate cyclohydrolase a bifunctional protein requiring dimerization for transformylase activity but not for cyclohydrolase activity. Journal of Biological Chemistry 276(11):7727-33.

Wolan, D.W., Cheong, C.G., Greasley, S.E., and Wilson, I.A. 2004. Structural insights into the human and avian IMP cyclohydrolase mechanism via crystal structures with the bound XMP inhibitor. Biochemistry 43(5):1171-83.

Wolan, D.W., Greasley, S.E., Beardsley, G.P., and Wilson, I.A. 2002. Structural insights into the avian AICAR transformylase mechanism. Biochemistry 41(52):15505-13.

Wolan, D.W., Greasley, S.E., Wall, M.J., Benkovic, S.J., and Wilson, I.A. 2003. Structure of avian AICAR transformylase with a multisubstrate adduct inhibitor β-DADF identifies the folate binding site. Biochemistry 42(37):10904-14.

Yao, J., Saenkham, P., Levy, J., Ibanez, F., Noroy, C., Mendoza, A., Huot, O., Meyer, D.F., and Tamborindeguy, C. 2016. Interactions "*Candidatus* Liberibacter solanacearum"—*Bactericera cockerelli*: haplotype effect on vector fitness and gene expression analyses. Frontiers in Cellular and Infection Microbiology 9(6):62.

# Chapter Four

# New approaches to express and purify functional *Ca*. L. solanacearum proteins

Part of the work presented in this Chapter has been published in:

Gilkes, J.M., Sheen, C.R., Frampton, R.A., Smith, G.R., and Dobson, R.C.J. 2019. The first purification of functional proteins from the unculturable, genome reduced, bottlenecked alpha-proteobacterium *Candidatus* Liberibacter solanacearum. Phytopathology 109(7):1141-1148.

## Preface

Chapter Three demonstrated that soluble, folded *C*Lso adenyltransferase/IMP cyclohydrolase can be produced *via* the co-expression of chaperone proteins. This begged the question as to whether adenyltransferase/IMP cyclohydrolase was unique in its difficulty in acquiring soluble-folded protein, or whether this was a characteristic of genome reduction, causing proteins from *C*Lso to aggregate or misfold. I was interested in how *C*Lso proteins fold *in vivo*, since understanding this would inform *in vitro* expression and purification. An understanding of those mechanisms could provide a solution to the expression and purification of *C*Lso proteins *in vitro* by potentially mimicking the cellular environment of *C*Lso for protein folding. Here, an introduction into the bottleneck effect is described as it is believed that this process is responsible for the destabilisation of protein structure in intracellular bacteria. Additionally, chaperone expression in the *C*Lso genome is discussed.

## 4.1 Introduction

### 4.1.1 The bottleneck effect

*C*Lso has a reduced genome of 1.26 Mbp (Lin et al. 2011). Genomes of less than 1 Mbp are found in host-dependent microorganisms, reflecting that the richness of metabolites provided by the host allows for the loss of genes that were previously deemed essential (McCutcheon and Moran 2012). Genes maintained for either parasitic or mutualistic benefit tend to include those for cellular processes or interactions with the host (McCutcheon and Moran 2012), such as the co-synthesis of essential amino acids, vitamins and carotenoids (Shigenobu et al. 2000; McCutcheon and Moran 2012; Sloan and Moran 2012). Additionally, reduced genomes generally lack the genes for DNA repair and have a bias towards A and T residues (Ochamn and Kelkar 2013). This, in turn, results in a mutational bias towards hydrophobic amino acid residues in the encoded proteins, increasing the likelihood of protein aggregation (Ochman and Kelkar 2013).

Although all genomes are subject to mutation *via* for example deletions, insertions or substitutions, most mutations are either neutral, or selected against and removed from the population if there is a negative effect on fitness (McCutcheon and Moran 2012). Asexual

reproduction of intracellular bacteria and vertical transmission through the host affects the endosymbiont population as it experiences frequent population “bottlenecks” (Wernegreen 2002; McCutcheon and Moran 2012). Liberibacter species experience a reduction in the number of bacteria that are transmitted to the subsequent insect generation, resulting in a “bottleneck” in sustaining symbionts through successive generations. However, horizontal transmission, *via* the insects feeding on the host plant, off-sets some of the consequences of the restrictions on vertical transmission (Hosokawa et al. 2007; Mira and Moran 2002). Workneh et al. (2016) released *C*Lso positive pysllids in covered field plots, however, the proportion of psyllids that tested positive for *C*Lso declined over time. It was proposed that this decline in *C*Lso positive psyllids over time could be the result of inefficient vertical transmission of *C*Lso (Mira and Moran 2002; Workneh et al. 2016). Psyllids lay multiple eggs at one time and thus it is possible that some eggs may harbour the symbiont, whilst others may not (Workneh et al. 2016). Successive bottlenecks allow deleterious mutations to become fixed through genetic drift (Wernegreen 2002; McCutcheon and Moran 2012; Wernegreen 2015). Such mutations generally reduce the thermodynamic stability of expressed proteins (van Ham et al. 2003) resulting in destabilisation of protein secondary structure and making the proteins more prone to misfolding and aggregation (Wernegreen and Moran 1999; Mira and Moran 2002).

Liberibacter species have been shown to experience ‘bottlenecks’ from vertical transmission (from vector mother to offspring) and due to horizontal transmission (vector acquisition of the pathogen during feeding on an infected hosts) (Nachappa et al. 2014). Bottlenecks in pathogen transmission, along with the lack of DNA repair mechanisms and an A/T bias, could result in the destabilisation of protein secondary structure for *C*Lso proteins not limited to adenyltransferase/IMP cyclohydrolase.

### 4.1.2 Increased chaperone expression in the *C*Lso genome

A series of studies have revealed that intracellular bacteria have increased chaperone expression in order to compensate for deleterious mutations that affect protein folding and stability (Moran 1996; Fares et al. 2002: van Ham 2003; Fares et al. 2004; McCutcheon and Moran 2012; Wernegreen 2015). Nachappa et al. (2012) found that the chaperone proteins GroEL, DnaK and GrpE were highly expressed in *C*Lso infected psyllids. Genetic drift should

lead to a decline in genomic fitness and eventual extinction (Wernegreen 2002; McCutcheon and Moran 2012; Wernegreen 2015). However, the persistence of endosymbiotic relationships implies a mechanism that buffers against the effect of deleterious genetic variation, preventing the organisms from experiencing a fitness decline (Fares et al. 2004). Chaperones are abundantly expressed and are conserved among intracellular bacteria (van Ham 2003, McCutcheon and Moran 2012). Shotgun proteomics studies have found that GroEL and DnaK are the most abundant proteins in *Ca*. Hodgkinia cicadicola and *Ca*. Sulcia muelleri (McCutcheon and Moran 2012). Van Ham et al. (2003) computationally demonstrated thermodynamic instability of *Buchnera* proteins and determined the proteins would have reduced and slower protein folding efficiency, misfolding and aggregation, or instability in native conformation. In contrast, the DnaK chaperonin of *Buchnera* displayed thermodynamic stability similar to free living bacteria (van Ham et al. 2003). In addition, Fares et al. (2002) demonstrated that GroEL overproduction in *Escherichia coli* compensates for deleterious mutations that accumulated in a simulated process of vertical transmission.

### 4.1.3 Overview of the Chapter

*C*Lso proteins are prone to misfolding and aggregation, as demonstrated by the many efforts by us and others, to obtain soluble folded proteins, *e.g.* serralysin (Ravindran et al. 2017) and adenyltransferase/IMP cyclohydrolase. However, to determine whether *C*Lso proteins are generally unstable, a variety of proteins would need to be characterised in order to verify that protein destabilisation is a feature of the *C*Lso genome. Furthermore, an increase in chaperone expression in *C*Lso is believed to occur in order to overcome the effects of successive bottlenecks in *C*Lso and the consequences of genome reduction.

As discussed in Chapter One, Section 1.8, there are no studies that investigate *C*Lso protein structure; many studies involve protein denaturation and refolding, which requires extensive protein characterisation for the refolded protein to ensure it has retained its native fold and oligomeric state for future functional and structural characterisation (Liu et al. 2017; Ravindran et al. 2018). Thus, the aim of this research was to develop a general method for the expression and purification of functional *C*Lso proteins in their native state, without the need for subsequent denaturation and refolding.

I reasoned that since chaperone proteins are highly overexpressed in the *C*Lso genome, they could be required for *C*Lso protein folding, thus an *in vitro* approach to mimic cellular conditions via the co-expression of chaperones would allow the isolation of correctly folded and functional proteins from the organism. To test this, and to determine whether protein destabilisation occurred for multiple *C*Lso proteins, chaperone co-expression was investigated with: dihydrodipicolinate synthase (DHDPS, the rate limiting step in *L*-lysine biosynthesis and potential drug target) and pyruvate kinase (PK, a core metabolic enzyme involved in glycolysis). Additionally, the following putative pathogenicity *C*Lso protein were also investigated: dTDP-4-dehydrorhamnose 3,5-epimerase (*rmlC*, required for the biosynthesis of dTDP-L-rhamnose and is a high confidence drug target), dTDP-4-dehydrorhamnose reductase (*rmlD*, required for dTDP-L-rhamnose synthesis and formation of extracellular polysaccharide matrix), and serralysin (metalloprotease known to be a bacterial virulence factor).

Further purification and characterisation was conducted with DHDPS and PK as other bacterial orthologues have been studied in our lab and were well characterised. Thus, resources and expertise were available to test activity and to determine the differences between previously investigated DHDPS, PK and the *C*Lso counterparts.

## 4.2 Results and Discussion

### 4.2.1 Protein expression in a standard *E. coli* system

In order to determine whether other proteins in the *C*Lso genome were prone to misfolding and aggregation, the proteins of interest were initially overexpressed in *E. coli* Tuner(DE)pLacI cells from the pOPINF plasmid. However, all expressed proteins were insoluble (Figure 4.1, refer to Chapter Eight for methodology). SDS-PAGE analysis showed that *C*Lso DHDPS was overexpressed as a ~35 kDa insoluble protein (Figure 4.1A), consistent with its predicted molecular mass of 32.6 kDa. PK was overexpressed as a 53 kDa insoluble protein (Figure 4.1B), also consistent with its predicted molecular mass of 52.8 kDa. Dehydrorhamnose epimerase and dehydrorhamnose reductase were over expressed as an ~21 kDa and ~30 kDa insoluble protein respectively (Figure 4.1C) consistent with the predicted molecular masses of 23.6 kDa for dehydrorhamnose epimerase and 31.7 kDa for dehydrorhamnose reductase. Serralysin

was overexpressed as an ~55 kDa insoluble proteins (Figure 4.1D) which correlated with the expected molecular mass of 57 kDa.

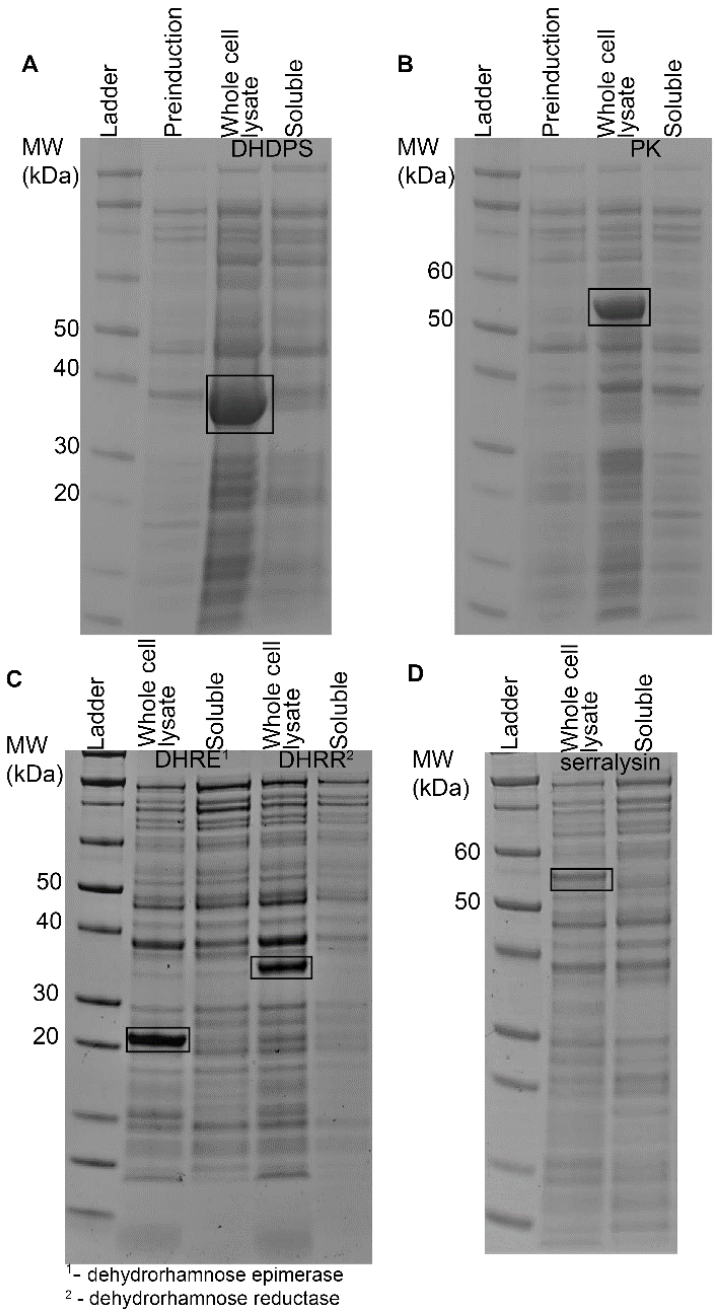



**Figure 4.1:** SDS-PAGE analysis of expressed *C*Lso proteins that were insoluble in solution. **A** shows DHDPS, **B** is PK, whilst **C** shows dehydrorhamnose epimerase and dehydrorhamnose reductase, and **D** is serralysin.

Recombinant protein expression in *E. coli* produced active enzymes that were soluble and easily purified for DHDPS from for example *Vitis vinifera* (Atkinson et al. 2011), *Bacillus anthracis* (Domigan et al. 2009), and *Mycobacterium tuberculosis* (Evans et al. 2011), and PK

from *Salmonella typhimurium* (Olalla and Garrido-Pertierra 1987), *Leishmania mexicana* (Tulloch et al. 2008) and *Toxoplasma gondii* (Bakszt et al. 2010). In addition, dehydrorhamnose epimerase and dehydrorhamnose reductase have also been recombinantly expressed and purified in *E. coli* from *Salmonella enterica* (Giroud et al. 1999; Graninger et al. 1999). As a result, it is surprising that a similar approach to the expression and purification of these proteins from *C*Lso produced insoluble protein. Like the adenyltransferase/IMP cyclohydrolase, proteins from *C*Lso show an inherent disposition to persist as insoluble proteins in solution.

#### *4.2.1.1 Does protein hydrophobicity or gene A/ T composition account for insoluble CLso proteins?*

To determine whether gene hydrophobicity or A/T bias could account for insoluble *C*Lso proteins, the genes of interest were compared to the *E. coli* and other highly reduced genome gene homologs (*B. aphidicola* and *Ca*. Carsonella rudii). For all the *C*Lso proteins investigated, there was no significant increase in grand average hydropathicity (GRAVY) as A – T compositon increased for any of the proteins (relative to *E. coli*) (Table 4.1). For *dapA* encoding DHDPS, the A – T composition of the gene increased with decreasing genome size, there was also an increase in hydrophobicity of the protein (Table 4.1). This correlates with the theory proposed by Ochman and Kelkar 2013 that an A – T mutational bias could result in proteins with more hydrophobic residues. However, an increase in hydrophobicity was not observed for the other proteins.

**Table 4.1:** Comparison of *C*Lso A – T gene composition and hydrophobicity of the encoded proteins with *Escherichia coli* and other reduced genome organisms

| | | | | | Protein (gene) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Organism** | **Accession Number** | **Genome AT content (%)** | **Genome Size** | | **DHDPS (*dapA*)** | **PK (*pykF*)** | **ATIC (*purH*)** | **DHRE (*rmlC*)** | **DHRR (*rmlD*)** | **Serralysin (*ZnMc*)** |
| *Escherichia coli* | NC_011750.1 | 50.6 | 5.13 Mbp[1] | GRAVY[1] | 0.057 | -0.031 | -0.126 | -0.346 | -0.108 | - |
| | | | | A – T % | 48.2 | 50.2 | 46 | 67 | 53.4 | - |
| *Ca.* L. solanacearum | NZ_JMTK00000000.1 | 64.8 | 1.26 Mbp | GRAVY | 0.06 | -0.029 | -0.113 | -0.287 | -0.076 | -0.477 |
| | | | | A – T % | 61.3 | 58.7 | 46 | 62 | 46.5 | 47.5 |
| *Buchnera aphidicola* | NC_002528 | 74.74 | 640 Kbp | GRAVY | 0.07 | -0.073 | -0.254 | - | - | - |
| | | | | A – T % | 67.9 | 69 | 72.9 | - | - | -- |
| *Ca*. Carsonella rudii | NC_018417 | 85.3 | 155 Kb | GRAVY | 0.168 | - | - | - | - | - |
| | | | | A – T % | 85.5 | - | - | - | - | - |

[1]GRAVY, (grand average hydrophobicity) - A positive value indicates more hydrophobic proteins whilst a negative value represents hydrophilic proteins

*4.2.1.2 CLso proteins are prone to aggregation*

To determine whether the *C*Lso proteins have a greater propensity for protein aggregation, AGGREGSCAN (Conchillo-Sole et al. 2007; de Groot et al. 2012) was used to predict aggregation-prone regions in the protein sequences of interest. AGGRESCAN is based on an aggregation-propensity scale for amino acids and assumes that short and specific sequence stretches modulate protein aggregation (Conchillo-Sole et al. 2007). The intrinsic aggregation properties for the proteins of interest were compared between *C*Lso, *E. coli*, *A. tumefaciens* and *B. aphidicola*. The average aggregation propensity of the sequence (Na4vSS) and the hot spot area (HAS) for each amino acid is shown in Figure 4.2. The DHDPS, PK and adenyltransferase/IMP cyclohydrolase (investigated in Chapter Three) of *B. aphidiola* and *C*Lso show a greater propensity for protein aggregation compared to *E. coli* and *A. tumefaciens* (Figure 4.2 ABC). Dehydrorhamnose reductase of *C*Lso also shows the protein to be prone to aggregation *in vitro*. Here, *B. aphidicola* proteins are shown to be prone to aggregation, consistent with the computational studies of Van Ham et al. (2003), and demonstrating that *B. aphidicola* proteins have a reduced and slower protein folding efficiency, a misfolding and aggregation propencity, or are unstable in their native conformation. Further, *B. aphidicola* has a significantly more reduced genome (640 kbp) than *C*Lso (1.26 Mbp), although the proteins for both organisms have higher aggregation propensities for DHDPS, PK and adenyltransferase/IMP cyclohydrolase, the results are consistent with an increase in protein instability as genome size decreases.

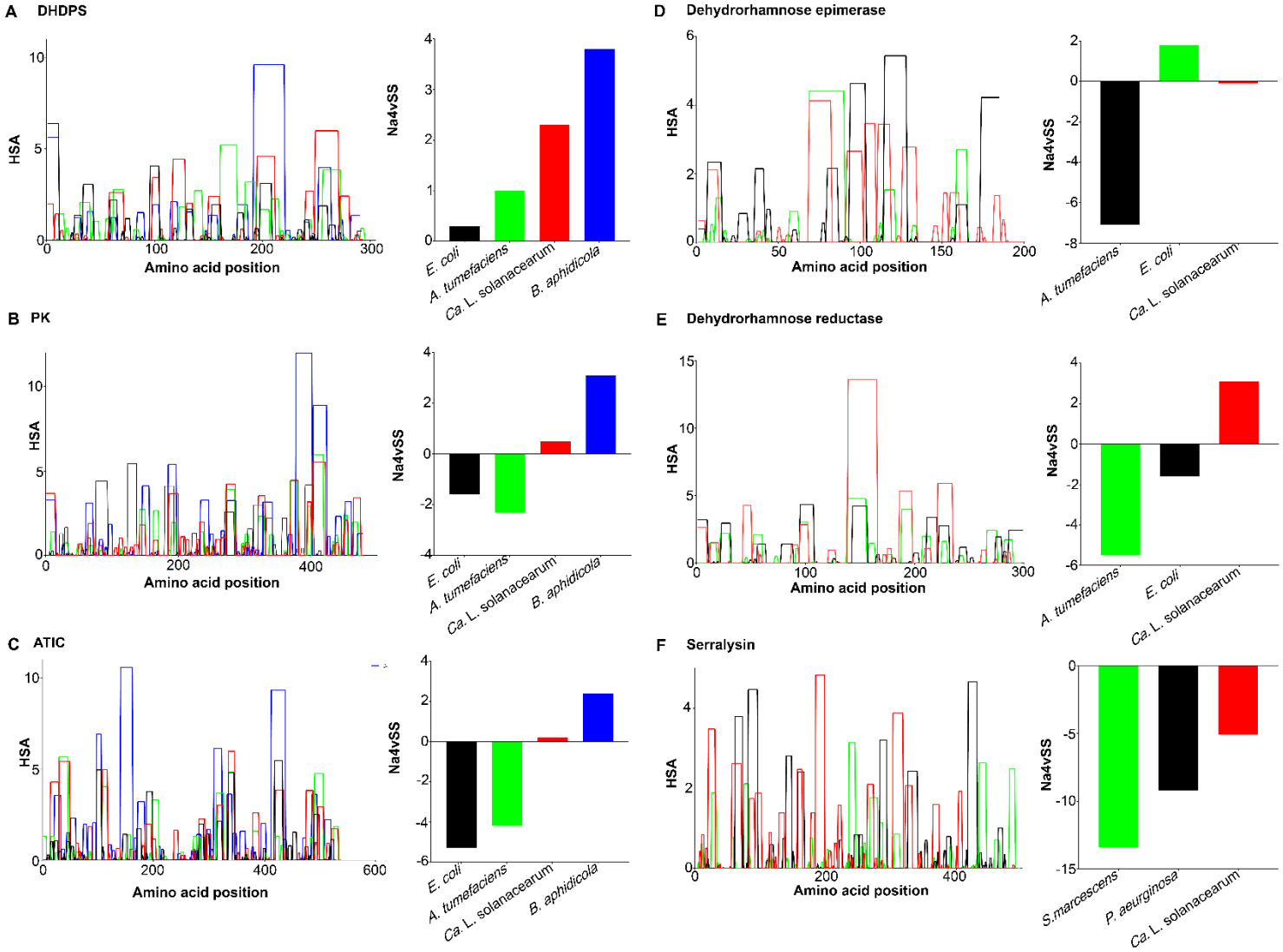



**Figure 4.2:** Average protein aggregation propensity predicted by AGGRESCAN. Included is the hot spot area (HSA) for each amino acid of the protein, high HAS values correlate with regions of the protein that are prone to aggregation.

### 4.2.2 Soluble *C*Lso protein when co-expressed with chaperones

I reasoned that, since bottlenecked bacteria with reduced genomes overexpress chaperones to compensate for protein instability, co-expression with chaperones might help express soluble *C*Lso proteins. As such, I expressed each protein with different combinations of *E. coli* chaperone proteins, to mimic the high expression levels of chaperones observed in *C*Lso *in vivo* (Table 4.2).

**Table 4.2:** Plasmids used for co-expression of chaperones with *C*Lso proteins.

| Plasmid | Chaperone | Expression Host | Chaperone molecular weight (kDa) | Promoter | Inducer |
|---|---|---|---|---|---|
| pGRO7 | GroES<br>GroEL | *E.coli* BL21(DE3) | 10<br>60 | *araB* | L-arabinose |
| pKJE7 | dnaK<br>dnaJ<br>grpE | *E.coli* BL21(DE3) | 70<br>40<br>22 | *araB* | L-arabinose |
| pTf16 | Trigger factor | *E.coli* BL21(DE3) | 56 | *araB* | L-arabinose |

Consistent with this hypothesis, *C*Lso DHDPS was expressed as a soluble 32.6 kDa protein when co-expressed with GroEL and GroES from the pGRO7 plasmid (Figure 4.3A). Similarly, *C*Lso PK was also expressed as a soluble 53 kDa protein when co-expressed with GroEL/GroES (pGRO7) (Figure 4.2B), DnaK/DnaJ/GrpE (pKJE7) (Figure 4.2B), or trigger factor (pTF16) (Figure 4.3B). Interestingly dehydrorhamnose epimerase, dehydrorhamnose reductase and serralysin only soluble when co-expressed with the DnaK/DnaJ/GrpE chaperone proteins from the pKJE7 plasmid (Figure 4.3C).

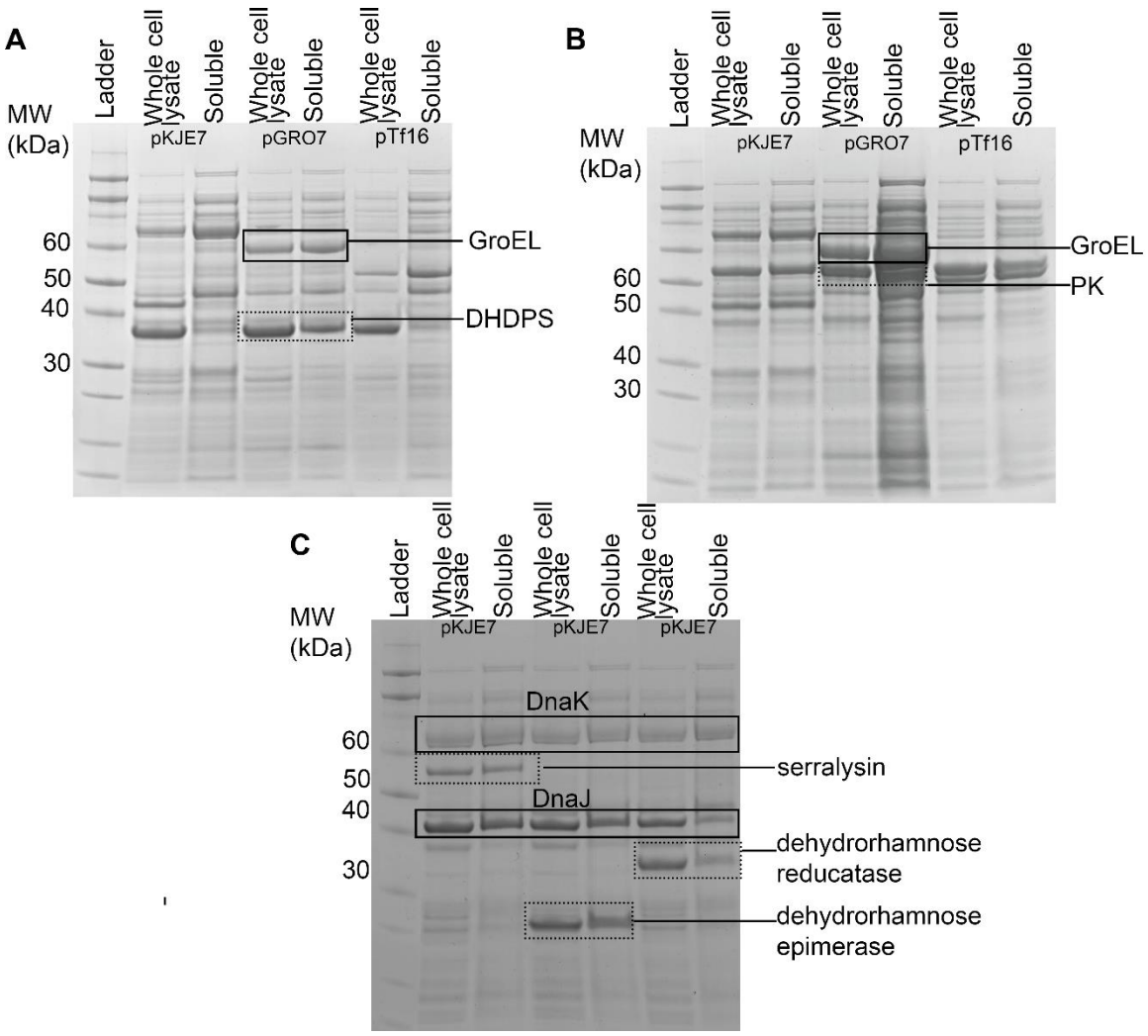



**Figure 4.3:** Co-expression of *C*Lso proteins with *E. coli* chaperone proteins. **A** and **B** show the expression of soluble DHDPS and PK with the co-expression of GroEL and GroES chaperone proteins. **C** shows the soluble expression of serralysin, dehydrorhamnose epimerase and dehydrorhamnose reducatase with the DnaK, DnaJ and GrpE chaperone proteins.

Interestingly, transcriptomic analysis of infected psyllids determined that GroEL, DnaK and GrpE were highly expressed genes in *C*Lso (Nachappa et al. 2012). Consistent with this observation, the pGRO7 plasmid, with GroEL expression, and the pKJE7 plasmid, encoding GrpE, appear to provide conditions that are sufficiently similar to those found in *C*Lso *in vivo* to allow *in vitro* production of soluble, functional protein.

The results demonstrate that soluble *C*Lso proteins can be expressed *via* the use of chaperone co-expression. The difficulty in obtaining soluble, stable *C*Lso proteins without the use of

chaperones is possibly caused by an inherent disposition of these proteins to aggregate or misfold. Both PK (glycolysis) and DHDPS (lysine biosynthesis) are core metabolic enzymes, whilst dehydrorhamnose epimerase, dehydrorhamnose reductase and serralysin were believed to be essential for *C*Lso pathogencity and survival. The need to overexpress these *C*Lso genes with chaperones to generate functional proteins, unlike their bacterial orthologues where only DHDPS requires chaperones (GroEL/ES), suggests they may have been affected by genetic drift, caused by the bottleneck effect and resulting reduced genome environment in which *C*Lso exists. It is likely that due to their importance, *C*Lso chaperone proteins are under strong purifying selection to buffer against the deleterious effects of accumulating mutations within the *C*Lso genome (*i.e.* poor protein structure/folding).

To further investigate the oligomeric structure and activity of the proteins, *C*Lso DHDPS, PK, dehydrorhamnose epimerase and dehydrorhamnose reductase were purified to homogeneity by a combination of immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) (Figure 4.4). IMAC was used as a purification step to avoid the co-purification of contaminating *E. coli* chromosomally-derived enzymes that were not $His_6$-tagged. *C*Lso DHDPS eluted from the SEC column at approximately 80 mL (Figure 4.4A), PK eluted at both 50 mL and 70 mL (Figure 4.4B). *C*Lso dehydrorhamnose epimerase eluted from the SEC column at approximately 82 mL (Figure 4.4C), whilst dehydrorhamnose reductase eluted at 95 mL (Figure 4.4D).

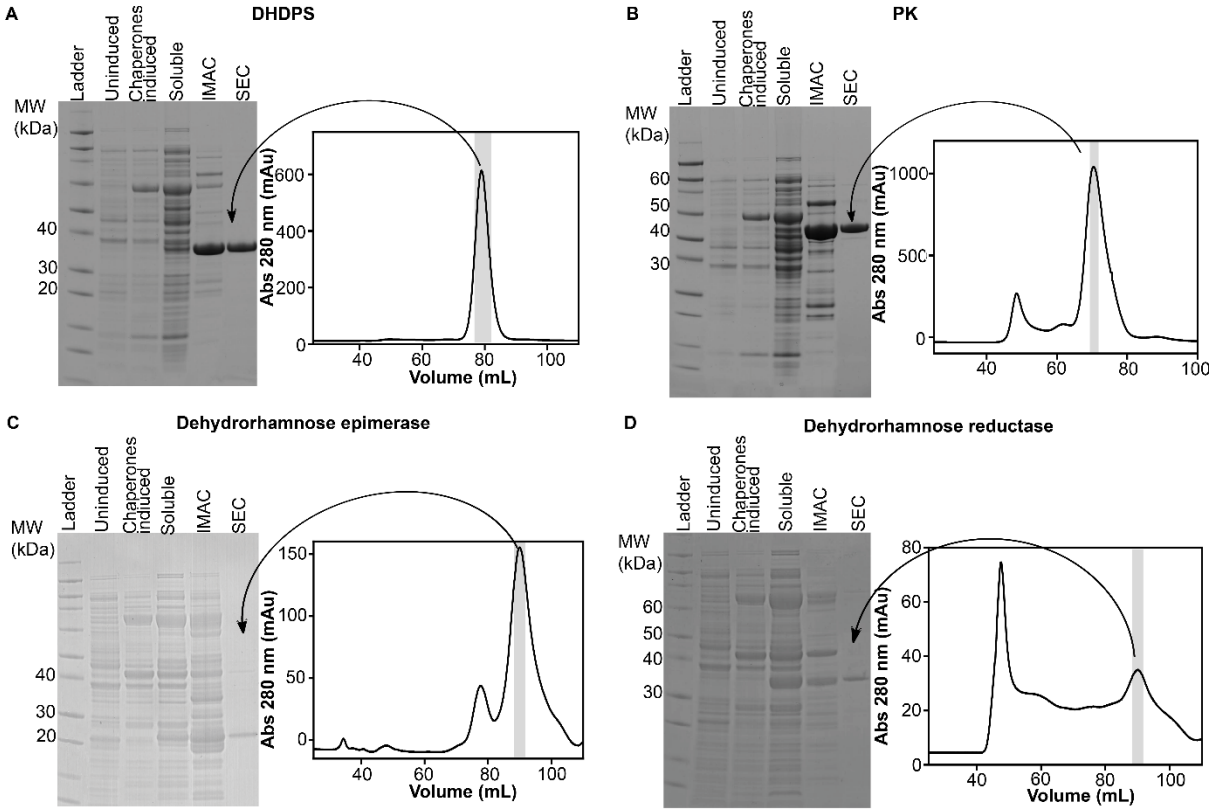



**Figure 4.4:** Purification of *C*Lso proteins after the co-expression of chaperone proteins. **A, B**, **C** and **D** show the purification of DHDPS, PK, dehydrorhamnose epimerase and dehydrorhamnose reductase, respectively, including the final SEC traces.

Previous purification strategies for *C*Lso proteins expressed in *E. coli* have used methods that denature the insoluble protein and then refold it (Liu et al. 2017; Ravindran et al. 2018). These methods can result in the protein assuming non-native secondary structures incurring loss of activity and protein misfolding or aggregation with the removal of the denaturant (Tsumoto et al. 2002). The research aimed to provide a method for the production of soluble and functional *C*Lso proteins. This Chapter has shown, for the first time, that chaperone co-expression approach can be successfully applied for the expression and purification of numerous *C*Lso proteins, this is a tremendous leap in Liberibacter biology and will aid in further understanding *C*Lso survival and pathogenicity.

### 4.2.3 Solution Structures

Although chaperone co-expression successfully expressed and purified soluble *C*Lso proteins, it is essential to characterise their structure and stability in solution to determine whether the proteins are correctly folded and stable in solution. From here, the Chapter will focus on the characterisation of DHDPS and PK.

First, I determined the thermal stability of the enzymes to ensure they were stable in the assay conditions. Differential scanning fluorimetry was used in order to determine the thermostability of *C*Lso DHDPS and PK. Both proteins were relatively stable, with DHDPS showing a melting temperature of 65.6°C and PK having a melting temperature of 56.4°C (Figure 4.5), giving confidence they would be stable in the activity assays (30°C).

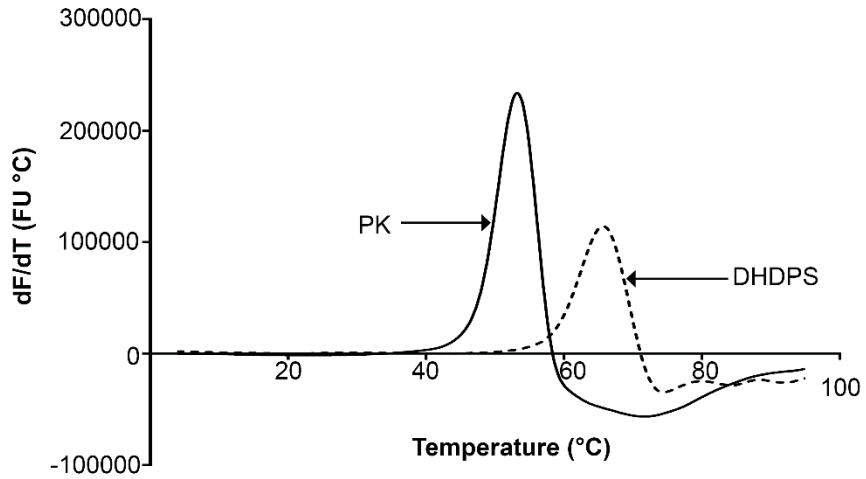



**Figure 4.5:** Thermostability profile indicating the unfolding temperature of DHDPS is 65.6°C and PK is 56.4°C.

Circular dichroism was used to test whether *C*Lso DHDPS and PK were folded and exhibited similar secondary structure to their *E. coli* counterparts (Figure 4.6). *C*Lso DHDPS (Figure 4.6A) has 28% α-helix, 24% β-structure, 20% β-turn and 28% unordered structure with a root mean square deviation (RMSD) of 0.029, whereas *E. coli* DHDPS has 17% α-helix, 28% β-structure, 25% β-turn and 30% unordered structure with a RMSD of 0.038. *C*Lso PK (Figure 4.6B) has 12% α-helix, 38% β-structure, 12% β-turn and 38% unordered structure with a RMSD of 0.060, whereas *E. coli* PK has 7% α-helix, 41% β-structure, 13% β-turn and 39% unordered structure with a RMSD of 0.032.

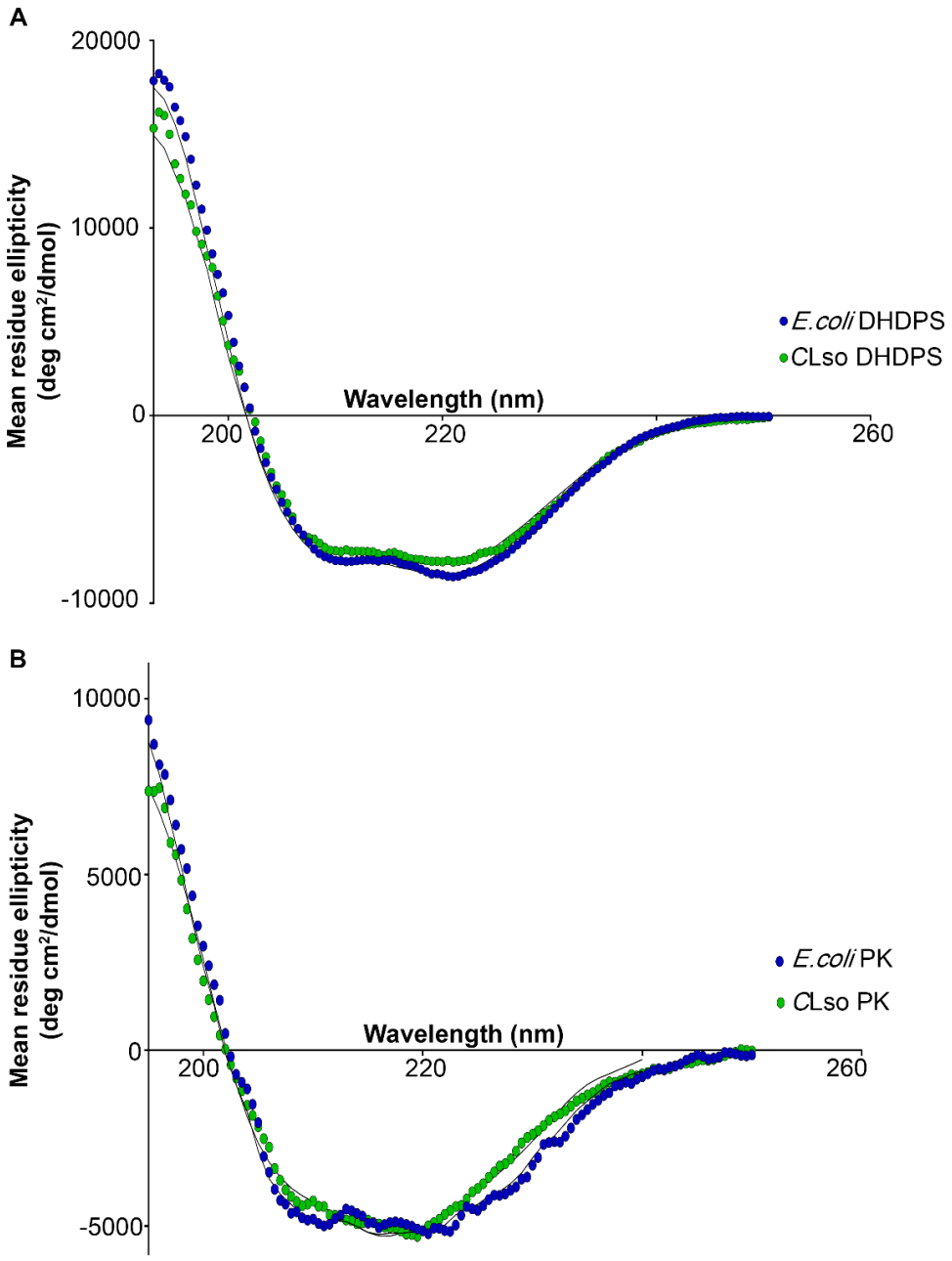



**Figure 4.6:** Circular dichroism spectroscopy. **A** is the spectrum for DHDPS and **B** is the spectrum for PK. Both show that the *C*Lso and *E. coli* enzymes are similarly folded. Raw data for DHDPS and PK from *C*Lso (green) and *E. coli* (blue) were fitted by nonlinear regression using the DICHROWEB software and the CDSSTR and CONTINLL algorithms.

*4.2.3.1 Quaternary structure as determined by analytical ultracentrifugation*

To determine the quaternary structure in solution for *C*Lso PK and DHDPS, sedimentation velocity experiments were conducted. The absorbance versus radial position profile for both proteins at concentrations of 0.2 mg/mL, 0.4 mg/mL and 0.6 mg/mL show a distinct sedimentation boundary profile consistent with the presence of a single species (Figure 4.7 for DHDPS and Figure 4.8 for PK).

The c(*M*) distributions for *C*Lso DHDPS shows a dominant, symmetrical peak, confirming that the protein largely exists as a single species in solution, without the presence of any aggregated protein (Figure 4.7). Continuous size-distribution analyses of the data determined that it had a standardized sedimentation coefficient of 6.8 S (Figure 4.7A). *C*Lso DHDPS data fitted to an apparent molar mass of 144 kDa, consistent with the theoretical tetrameric mass of 130 kDa. A single peak dominated at all concentrations, suggesting that the protein was not in a concentration dependent self-association. Both *C*Lso DHDPS and PK enzymes have the canonical tetrameric oligomeric structure in solution, consistent with other bacterial orthologues (Johnson and Deal 1970; Griffin et al. 2010; Muscroft-Taylor et al. 2010; Anastasiou et al. 2012).

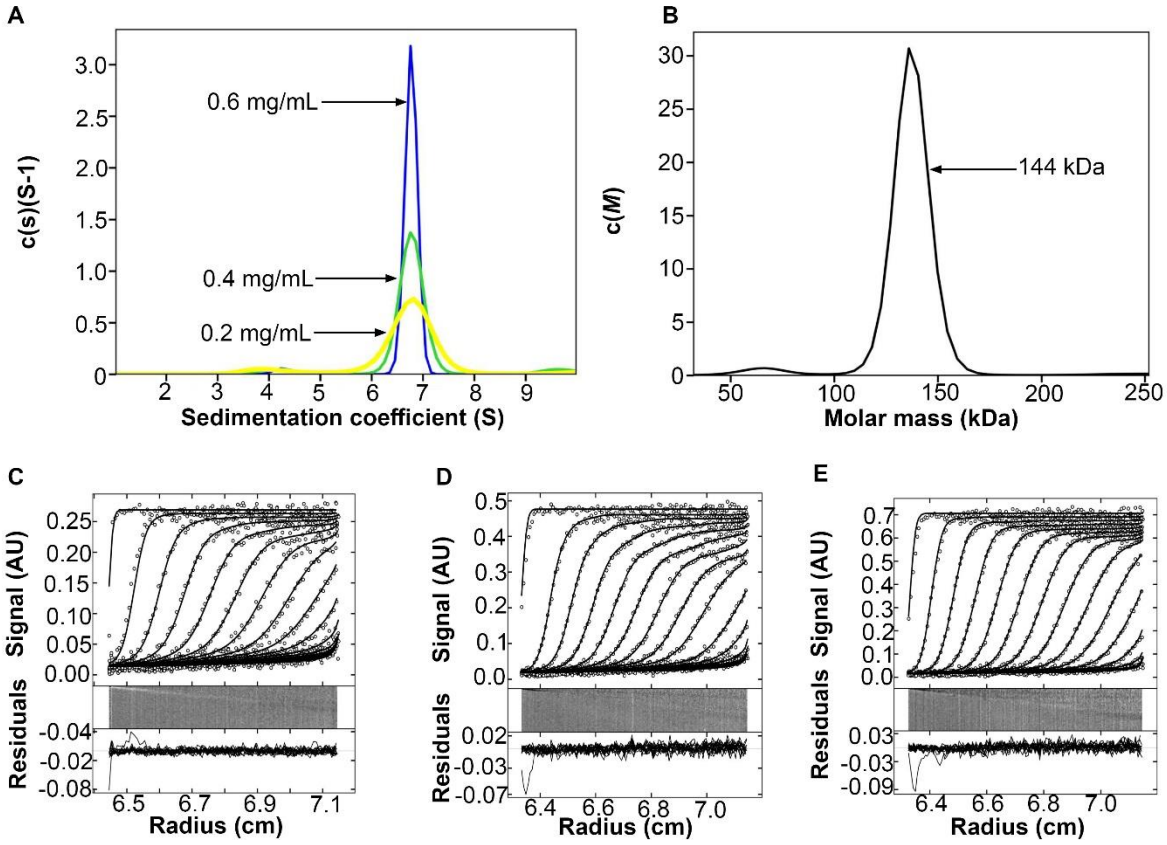



**Figure 4.7:** Sedimentation velocity analysis of *C*Lso DHDPS demonstrates a homogeneous sample with a mass of 144 kDa. Data were collected at 280 nm, 50 000 rpm and 20 °C. Data were fitted to a c(*s*) and c(*M*) model at a resolution of 300 and a confidence level of 0.95 using SEDFIT (Schuck, 2000). **A** is the c(*s*) model plotted as a function of sedimentation coefficient for 0.6 mg/mL, 0.4 mg/mL and 0.2 mg/mL. Data were fitted with an s value ranging between 1 S and 10 S. The fit resulted in a frictional ratio ($f/f_0$) of 1.91. **B** is the c(*M*) model plotted as a function of molecular mass (kDa). Data were fitted with a minimum mass of 20 kDa and a maximum mass of 250 kDa. **C**, **D** and **E** are the absorbance at 280 nm plotted as a function of radial position (cm) at 0.6 mg/mL, 0.4 mg/mL and 0.2 mg/mL respectively. The raw data are represented as open symbols (o) and overlaid with the non-linear least squares best fit. The residuals for this fit are shown above. Every third scan is shown.

The c(*M*) distributions for *C*Lso PK also shows a dominant, symmetrical peak, confirming that the protein largely exist as a single species in solution, without the presence of aggregated protein (Figure 4.8A). A continuous size-distribution analysis of the data from *C*Lso PK determined that it had a standardized sedimentation coefficient of 8.5 S (Figure 4.8A). The data fitted to apparent molar mass of 195 kDa, consistent with the theoretical tetrameric

mass of 213 kDa. This data is consistent with *E. coli* PK, that also shows a single tetrameric species of ~195 kDa (Zhu et al. 2010).

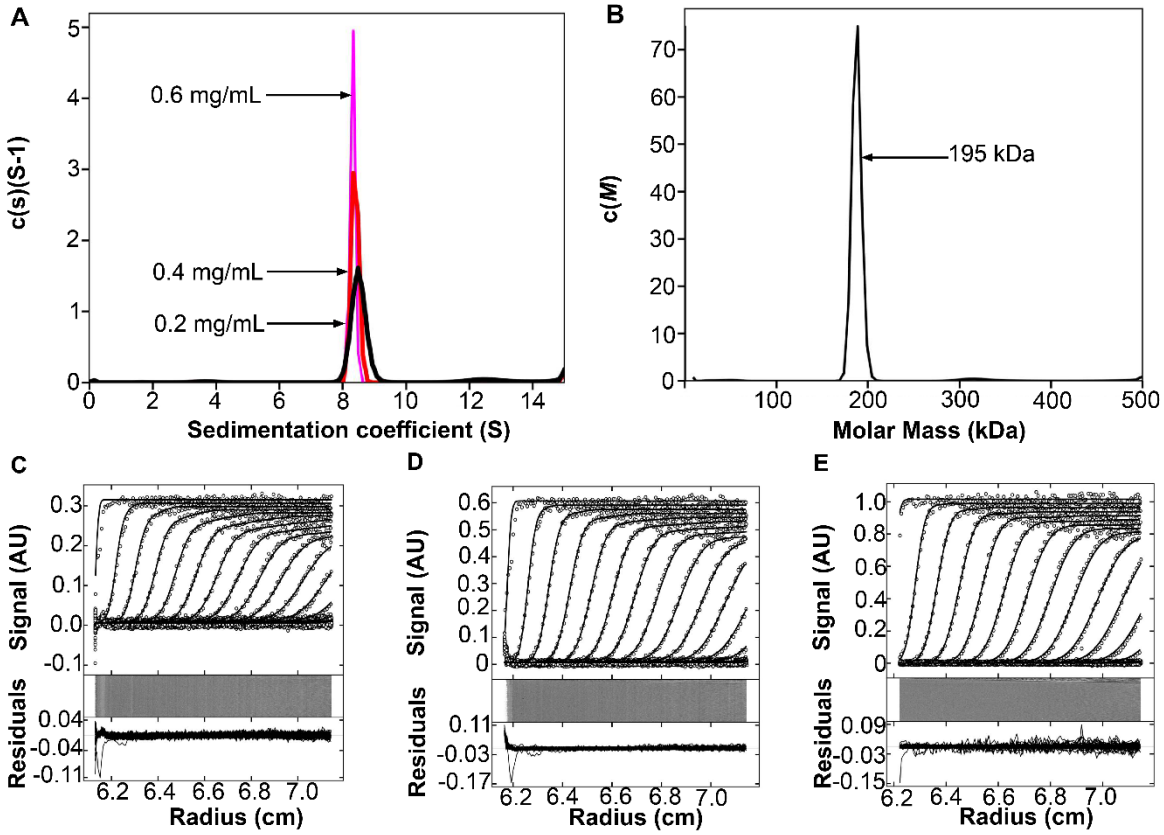



**Figure 4.8:** Sedimentation velocity analysis of *C*Lso PK demonstrates a homogeneous sample with a mass of 195 kDa. Data were collected at 280 nm, 50 000 rpm and 20 °C. Data were fitted to a c(*s*) and c(*M*) model at a resolution of 300 and a confidence level of 0.95 using SEDFIT (Schuck, 2000). **A** is the c(*s*) model plotted as a function of sedimentation coefficient for 0.6 mg/mL, 0.4 mg/mL and 0.2 mg/mL. Data were fitted with an s value ranging between 0 S and 14.5 S. The fit resulted in a frictional ratio (f/f0) of 1.2. **B** is The c(*M*) model plotted as a function of molecular mass (kDa). Data were fitted with a minimum mass of 20 kDa and a maximum mass of 500 kDa. **C**, **D** and **E** are the absorbance at 280 nm plotted as a function of radial position (cm) at 0.6 mg/mL, 0.4 mg/mL and 0.2 mg/mL respectively. The raw data are represented as open symbols (o) and overlaid with the non-linear least squares best fit. The residuals for this fit are shown above. Every third scan is shown.

### *4.2.3.2 Small angle X-ray Scattering (SAXS) analysis*

Small angle X-ray scattering was used to verify the solution structure of *C*Lso DHDPS and PK determined by sedimentation velocity experiments. Data quality was assessed by Guinier analysis of the SAXS scattering curve at low angles. The linearity of the Guinier plot for both

DHDPS (Figure 4.9A inset) and PK (Figure 4.10D inset), confirms that the scattering of both *C*Lso proteins is reliable and no sample aggregation and/or inter-particle interference is present (Putnam et al. 2007). Scattering data are presented as an intensity plot (Figure 4.9A and 4.10A for DHDPS and PK respectively).

The radius of gyration ($R_g$) determined by GNOM (Svergun 1992) for DHDPS was calculated as 31.08 which is in agreement with the Guinier analysis (Figure 4.8). Scattering data are presented as an intensity plot (Figure 4.9A). In an effort to determine whether the solution structure of *C*Lso DHDPS is consistent with known DHDPS structures, experimental scattering was compared to the theoretical generated from the atomic coordinates of *Agrobacterium tumefaciens* (PDB ID: 4I7U) using CRYSOL (Svergun et al. 1995). *A. tumefaciens* shows 60% sequence identity with *C*Lso DHDPS. The scatter for the *A. tumefaciens* DHDPS tetramer fit the *C*Lso DHDPS scatter with a $\chi^2$ value of 0.3 indicating a good fit to the data and further indicates that *C*Lso DHDPS exists as a tetramer in solution (Figure 4.9A). The maximum dimension of the scattering particle ($D_{max}$) is 89.36 for *C*Lso DHDPS (Figure 4.9C). This correlates with the calculated $D_{max}$ of 81.6 for the tetrameric *Agrobacterium tumefaciens* DHDPS. The molecular mass determined by the SAXS data was 114 kDa. The mass as determined by SAXS for DHDPS was smaller than the apparent molecular mass of 144 kDa derived from sedimentation velocity experiments, however, the low resolution of SAXS data which could account for the small discrepancy in molecular weight (Putnam et al. 2007). The Kratky plot shows a bell shaped plot for DHDPS (Figure 4.9B) which further indicates the protein is folded and has some flexibility in solution. This indicates that *C*Lso DHDPS is globular in shape and not elongated in solution. The *P(r)* plot shows a single peak that is symmetrical in shape, with no positively skewed tail at long distance.

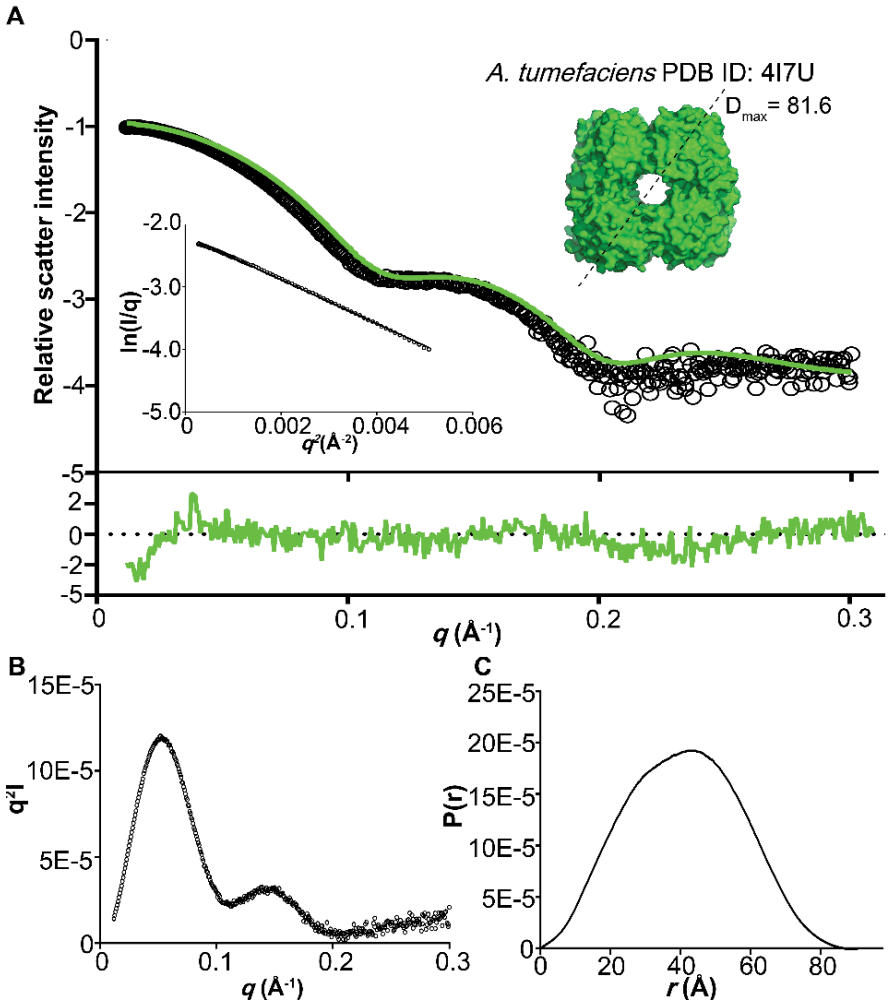



**Figure 4.9**: SAXS data for *C*Lso DHDPS. **A** is the experimental data (open circles, o) presented as an intensity plot, overlaid is the theoretical scattering profile for *A. tumefaciens* (PDB ID: 4I7U) ($\chi^2$ = 0.3, green fit line). The inset of **A** is the Guinier plot along with the surface structure of *A. tumefaciens* tetrameric DHDPS. **B** and C is the Kratky plot and *P(r)* plot respectively.

The radius of gyration ($R_g$) determined by GNOM (Svergun 1992) was calculated as 31.08 which is in agreement with the Guinier analysis, in addition the $R_g$ for PK was 44.84. In order to determine whether *C*Lso PK is consistent with known PK structures in solution, the experimental scatter was compared with the theoretical scattering generated from the atomic coordinates of *E. coli* PK (PDB ID: 1PKY) using CRYSOL (Svergun et al. 1995). The scatter for the *E. coli* PK tetramer fit the *C*Lso PK scatter with a $\chi^2$ value of 1.98 (Figure 4.10A). The poor fit of the *E. coli* PK tetramer could indicate a difference between *C*Lso PK in solution and the crystallised state of *E. coli* PK (Putnam et al. 2007), however, the sequence identity between *C*Lso and *E. coli* PK is only 40% and thus, the poor fit could indicate structural changes

in *C*Lso PK. This raises the question as to whether there are any significant structural changes in *C*Lso PK which could be the consequence of genome reduction and genetic drift (investigated in Chapter Five). The maximum dimension of the scattering particle ($D_{max}$) is 142.5 for PK (Figure 4.10C). This correlates with the calculated $D_{max}$ of 139.6 for the tetrameric *E. coli* PK. The molecular mass determined by the SAXS data was 190 kDa for PK. The molecular mass for PK determined by SAXS correlated with the mass determined from sedimentation velocity experiments. The Kratky plot shows a bell shaped plot for which further indicates the protein is folded and has some flexibility in solution. This indicates that *C*Lso PK is also globular in shape and not elongated in solution. The *P(r)* plot shows a single peak that is symmetrical in shape, with no positively skewed tail at long distance (Figure 4.10B). The structural parameters for both DHDPS and PK SAXS data are presented in Table 4.3.

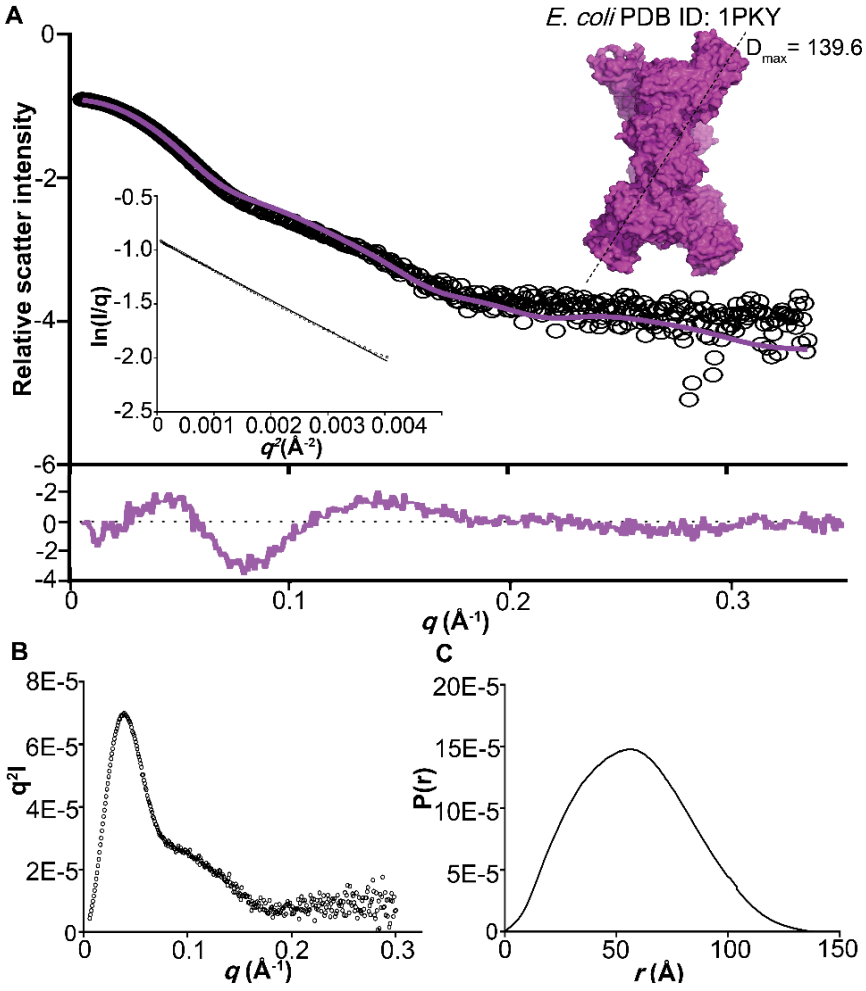



**Figure 4.10**: SAXS data for *C*Lso PK. **A** is the experimental data (open circles, o) presented as an intensity plot, overlaid is the theoretical scattering profile for *E. coli* (PDB ID: 1PKY) ($\chi^2$ = 1.98, purple fit line). The inset of **A** is the Guinier plot along with the surface structure of *E. coli* tetrameric PK. **B** and **C** is the Kratky plot and *P(r)* plot respectively.

**Table 4.3:** Data collection and analysis statistics from SAXS.

| **SAXS data analysis** | **DHDPS** | **PK** |
|---|---|---|
| $I(0)$ ($cm^{-1}$) (Guinier analysis) | 0.11 ± 0.0002 | 0.13 ± 0.00019 |
| $R_g$ (Å) (Guinier analysis) | 31.08 ± 0.30 | 44.57 ± 0.46 |
| $I(0)$ ($cm^{-1}$) ($P(r)$ analysis) | 0.11 | 0.12 |
| $R_g$ (Å) ($P(r)$ analysis) | 31.08 | 44.5 |
| $D_{max}$ (Å) | 89.36 | 142.5 |
| Porod volume ($Å^{-3}$) | 164 000 | 324 000 |
| Molar mass (Porod volume, kDa) | 96 470 | 190 588 |
| Molar mass (SAXSMoW2*, kDa) | 113 650 | 207 769 |
| Calculated tetramer mass from sequence (kDa) | 132 000 | 213 000 |
| **SAXS data collection parameters** | | |
| Instrument | Australian Synchrotron SAXS/WAXS beamline | |
| detector | PILATUS 1M (Dectris) | |
| wavelength (Å) | 1.0332 | |
| Maximum flux at sample | 8 x $10^{12}$ photons per second at 12 keV | |
| Camera length (mm) | 1600 | |
| Q range ($Å^{-1}$) | 0.006-0.5 | |
| Exposure time | Continuous 1 second frame measurements | |
| Sample configuration | SEC-SAXS with co-flow | |
| Sample temperature (°C) | 12 | |

***** http://saxs.ifsc.usp.br/

### 4.2.4 Activity assays

The final aim of the research was to determine whether co-expression of *C*Lso proteins with chaperones was able to produce functional proteins. Thus, activity assays for DHDPS and PK were conducted in order to determine whether the enzymes were functional. The specific activity was recorded in order to monitor the progress and success for each of purification step.

The specific activity of the DHDPS increased during purification (Table 4.4). The specific activity of DHDPS increased from 0.56 µmol/min/mg to 2.4 µmol/min/mg after SEC, demonstrating the protein had been purified and was active. The purification produced 9.6

mg of purified *C*Lso DHDPS per liter of culture. However, the total DHDPS activity decreased from 789 µmol/min in the crude lysate to 23 µmol/min after SEC (a yield of just 2.9%).

The specific activity of the PK increased during purification from 2.6 µmol/min/mg to 5.6 µmol/min/mg, demonstrating the protein had been purified and was active. The purification produced 2.9 mg of purified *C*Lso PK per liter of culture. However, the total PK activity decreased from 2 480 µmol/min in the crude lysate to 16 µmol/min after SEC (a yield of just 0.64%). Although the yields were low, there was sufficient enzyme for initial functional and structural characterization.

**Table 4.4:** Purification table for *C*Lso DHDPS and PK.

| Protein | Purification Step | Volume (mL) | [Protein] (mg) | Total Protein (mg) | Specific Activity (µmol/min/mg) | Total Activity (µmol/min) | Yield (%) | Purification Factor |
|---|---|---|---|---|---|---|---|---|
| DHDPS | Crude Lysate | 60 | 23.7 | 1419 | 0.56 | 789 | | |
| | IMAC | 10 | 2.2 | 22 | 0.51 | 11 | 1.5 | 0.92 |
| | SEC | 7.5 | 1.3 | 9.6 | 2.4 | 23 | 2.9 | 4.3 |
| PK | Crude Lysate | 60 | 16 | 954 | 2.6 | 2480 | | |
| | IMAC | 6 | 1.5 | 9.0 | 1.8 | 16 | 0.65 | 0.69 |
| | SEC | 6 | 0.48 | 2.9 | 5.6 | 16 | 0.64 | 2.2 |

The presence of L-lysine, a known allosteric inhibitor for other bacterial DHDPS enzymes (Karsten 1997; Griffin et al. 2012), significantly reduced the rate of reaction for DHDPS (Figure 4.11A dashed lines), suggesting that *C*Lso L-lysine biosynthesis is also allosterically regulated by L-lysine, as in other bacteria and plants. Additionally, PK demonstrated a rate of reaction of 0.085 ΔA340/min with 1.4 µg of the enzyme, the reaction doubled to 0.17 ΔA340/min with 2.9 µg, demonstrating that the rate of the reaction was dependent on the concentration of enzyme (Figure 4.11C). To ensure that the assay was working effectively, a positive control with the *E. coli* DHDPS and PK isozymes were included (Figure 4.11B and D). Interestingly the rates of reaction for both DHDPS and PK from *C*Lso were considerably slower compared to that of *E. coli*. The rate of reaction for *E. coli* DHDPS (4.4 ΔA340/min/0.3 µg) is increased ~32-fold compared to the rate of *C*Lso DHDPS (1.1 ΔA340/min/2.5 µg).

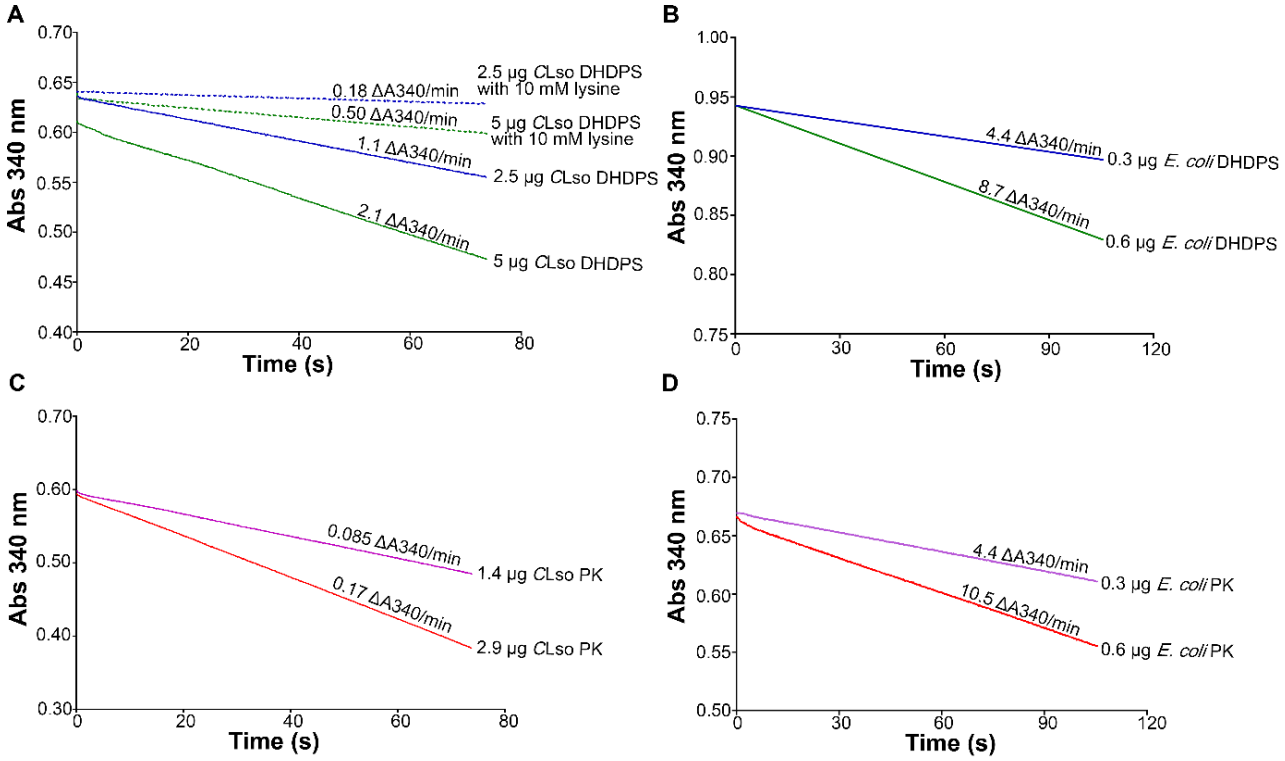



**Figure 4.11:** Activity assays for *C*Lso DHDPS and PK. **A** and **B** show the rates of reaction for *C*Lso and *E. coli* DHDPS respectively whilst **C** and **D** show the rates of reaction for *C*Lso and *E. coli* PK respectively. Rates of reaction are displayed for each plot. Both *C*Lso enzymes were active and DHDPS showed expected inhibition of DHDPS activity in the presence of lysine. Both DHDPS and PK displayed an increase in rate of reaction with increasing enzyme concentration.

The purification methods yielded sufficient quantities of protein, 9.6 mg of *C*Lso DHDPS and 2.9 mg of *C*Lso PK, from 1 L of culture. However, the specific activity of both *C*Lso DHDPS and PK were significantly lower than their *E. coli* counterparts. The specific activity of *C*Lso DHDPS was 2.4 µmol/min/mg with a yield of 2.9%, these values are lower than the recorded specific activity of 108 µmol/min/mg and yield of 39% for *E. coli* after purification (Dobson 2003). Similarly, the *C*Lso PK specific activity of 5.6 µmol/min/mg, with a yield of 0.64%, was considerably lower than the recorded specific activity for purified *E. coli* PK of 270 µmol/min/mg and 41% yield (Zhu 2008). This could reflect a loss of function during purification, which may be the result of the *C*Lso enzymes being less stable than their *E. coli* counterparts, demonstrated by the increase in *C*Lso aggregation propensity (Section 4.3.1.2). Additionally, a loss of function could also be the result of the crude lysate contains allosteric effectors not present in the subsequent purification steps.

Previous characterisation of *C*Lso proteins under denaturing conditions produced either non-functional protein or insoluble protein (Liu et al. 2017; Ravindran et al. 2018). In contrast, we demonstrated that co-expression of *C*Lso DHDPS and PK with chaperones resulted in active, tetrameric enzymes. Utilisation of chaperone co-expression circumvents the problems of protein refolding, because the formation of undesirable protein intermediate states is avoided. The activity of both *C*Lso DHDPS and PK corroborates that these proteins are in their native conformation. Furthermore, analytical ultracentrifugation determined both enzymes were present as tetramers in solution. Development of a method to obtain correctly folded, functional *C*Lso proteins provides a platform for further investigation into proteins that could play a key role in pathogenicity and understanding the requirements for survival of this genome-reduced organism.

## 4.3 Summary

This Chapter describes an applicable method for the expression and purification of numerous proteins from *C*Lso *via* the use of chaperone co-expression. This is a major advancement in *C*Lso biology. Previously, proteins from the organism were only able to be isolated denaturation and refolding. Although a number of studies note the difficulty in *C*Lso protein expression, there has been no hypothesis as to why *C*Lso protein expression has been such a challenge to date. Here, I propose that the intracellular lifestyle of the organism, results in a

bottleneck effect which allows deleterious mutations to become fixed through genetic drift. This along with the organisms' lack of DNA repair mechanisms has resulted in a destabilisation of protein secondary structure in *C*Lso proteins in which the overexpression of *C*Lso chaperone proteins aids in overcoming this effect. This explains why *C*Lso proteins are insoluble in a standard *E. coli* expression system and consequently require chaperone co-expression as this mimics the cellular environment of the organism.

The approach described was applicable for the expression and purification of the soluble *C*Lso proteins: DHDPS, PK, dehydrorhamnose epimerase, dehydrorhamnose reductase and serralysin. Characterisation of DHDPS and PK determined that both proteins were stable and folded in solution. In addition, solution structures determined via sedimentation velocity analysis and small angle X-ray scattering were consistent with both proteins retaining the canonical tetrameric structure seen with other bacterial orthologues.

Excitingly, both *C*Lso DHDPS and PK were shown to be active. This is the first time catalytic activity has been demonstrated for any protein from *C*Lso. Both *C*Lso DHDPS and PK demonstrated much lower rates of reaction compared to the *E. coli* enzyme. This raised interest as to whether there were any changes in substrate binding, affinity for the substrates or structural changes in the active site of each of the proteins. Thus, with a chaperone co-expression approach demonstrating the ability to obtain functional correctly folded *C*Lso proteins, further insights into protein structure can be undertaken with the aim in determining whether the proteins proposed in Chapter One could be potential targets for rationale antibacterial design.

## 4.4 References

Atkinson, S., Dogovski, C., Downton, M. T., Pearce, F.G., Reboul, C.F., Buckle, A.M., Gerrard, J.A., Dobson, R.C.J., Wagner, J., and Perugini, M.A. 2012. Crystal, solution and *in silico* structural studies of dihydrodipicolinate synthase from the common grapevine. PLoS One 7(6): e38318.

Atkinson, S.C., Dogovski, C., Downton, M.T., Czabotar, P.E., Dobson, R.C.J., Gerrard, J.A., Wagner, J., and Perugini, M.A. 2013. Structural, kinetic and computational investigation of *Vitis vinifera* DHDPS reveals new insight into mechanism of lysine mediated allosteric inhibition. Plant Molecular Biology 81(4-5): 431–446.

Atkinson, S.C., Dogovski, C., Newman, J., Dobson, R.C.J., and Perugini, M.A. 2011. Cloning, expression, purification and crystallization of dihydrodipicolinate synthase from the grapevine *Vitis vinifera*. Acta Crystallographica Section F 67(12): 1537–1541.

Bakszt, R., Wernimont, A., Allali-Hassani, A., Wai Mok, M., Hills, T., Hui, R., and Pizarro, J.C. 2010. The crystal structure of *Toxoplasma gondii* pyruvate kinase 1. PLos One 5(9): e12736.

Berrow, N.N., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I.,and Owens, R.J. 2007. A versatile ligation independent cloning method suitable for high throughput expression screening applications. Nucleic Acids Research 35(6):e45.

Blickling, S., Beisel, H.G., Bozic, D., Knablein, J., Laber, B., and Huber, R. 1997. Structure of dihydrodipicolinate synthase of *Nicotiana sylvestris* reveals novel quaternary structure. Journal of Molecular Biology 274(4): 608–621.

Boughton, B.A., Griffin, M.D.W., O'Donnell, P.A., Dobson, R.C.J., Perugini, M.A., Gerrard, J.A., and Hutton, C.A. 2008. Irreversible inhibition of dihydrodipicolinate synthase by 4-oxo-heptenedioic acid analogues. Bioorganic and Medicinal Chemistry 16(23): 9975–9983.

Boughton, B.A., Dobson, R.C.J., Gerrard, J.A., and Hutton, C.A. 2008b. Conformationally constrained diketopimelic acid analogues as inhibitors of dihydrodipicolinate synthase. Bioorganic & Medicinal Chemistry Letters 18(2): 460–463.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 7(72): 248–54.

Brown, P.H., and Schuck, P. 2008. A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifugation. Computer Physics Communications 178(2): 105–120.

Bucher, T., and Pfleiderer, G. 1955. Pyruvate kinase from muscle. In Colowick, S. P. and Kaplan, N. O. (eds), *Methods in Enzymology*. Academic Press, London, Vol. 1, pp. 435–440.

Buchman, J.L., Fisher, T.W., Sengoda, V.G., and Munyaneza, J.E. 2012. Zebra chip progression: from inoculation of potato plants with Liberibacter to development of disease symptoms in tubers. American Journal of Potato Research 89: 159–168.

Compton, L.A., and Johnson, W.C. 1986. Analysis of protein circular dichroism spectra for secondary structure using a simple matric multiplication. Analytical Biochemistry 155:155-157.

Crosslin, J.M., Munyaneza, J.E., Brown, J.K., and Liefting, L.W. 2010. Potato zebra chip disease; a phytopathological tale. Plant Health Progress 11(1): 33.

Dobson, R.C.J. 2003. Investigating the catalytic and regulatory mechanisms of dihydrodipicolinate synthase. PhD Thesis, University of Canterbury.

Dobson, R.C.J., Griffin, M.D.W., Devenish, S.R.A., Pearce, F.G., Hutton, C.A., Gerrard, J.A. Jameson, G.B., and Perugini, M.A. 2008.Conserved main-chain peptide distortions: A proposed role for Ile203 in catalysis by dihydrodipicolinate synthase. Protein Science 17(12): 2080–2090.

Dobson, R.C.J., Griffin, M.D.W., Roberts, S.J., and Gerrard, J.A. 2004a. Dihydrodipicolinate synthase (DHDPS) from *Escherichia coli* displays partial mixed inhibition with respect to its first substrate, pyruvate. Biochimie 86(4-5): 311–315.

Dobson, R.C.J., Valegard, K., and Gerrard, J.A. 2004b. The crystal structure of three site-directed mutants of *Escherichia coli* dihydrodipicolinate synthase: further evidence for a catalytic triad. Journal of Molecular Biology 338(2): 329–339.

Dobson, R.C.J., Griffin, M.D.W., Jameson, G.B.., and Gerrard, J.A. 2005. The crystal structures of native and (S)-lysine-bound dihydrodipicolinate synthase from *Escherichia coli* with improved resolution show new features of biological significance. Acta Crystallographica Section D D61: 1116-1124.

Dogovski, C., Atkinson, S.C., Dommaraju, S.R., Hor, L, Dobson, R.C.J., Hutton, C.A., and Gerrard, J.A. 2009. Lysine biosynthesis in bacteria: an unchartered pathway for novel antibiotic. Encyclopedia of life support systems 11: 116–136.

Donovan, K.A., Zhu, S., Liuni,, P., Kessans, S.A., Wilsons, D.J., and Dobson, R.C.J. 2016. Conformational dynamics and allostery in pyruvate kinase. Journal of Biological Chemistry 291(17): 9244–9256.

Domigan, L.J., Scally, S.W., Fogg, M.J., Hutton, C.A., Perugini, M.A., Dobson, R.C.J., Muscroft-Taylor, A.C., Gerrard, J.A., and Devenish, S.R.A. 2009. Characterisation of dihydrodipicolinate synthase (DHDPS) from *Bacillus anthracis*. Biochimica et Biophysica Acta 1794(19): 1510–1516.

Duan, Y., Zhou, L., Hall, D.G., Li, W., Doddapaneni, H., Lin, H., Vahling, C.M., Gabriel, D.W., Williams, K.P., Dickerman, A., Shun, Y., and Gottwald, T. 2009. Complete genome

sequence of citrus huanglongbing bacterium, 'Candidatus Liberibacter asiaticus' obtained through metagenomics. Molecular Plant-Microbe Interactions 22(8): 1011-1020.

Evans, G., Schuldt, L., Griffin, M.D., Devenish, S.R.A, Pearce, F.G.., Perugini, M.A., Dobson, R.C.J., Jameson, G.B., Weiss, M.S., and Gerrard, J.A. 2011. A tetrameric structure is not essential for activity in dihydrodipicolinate synthase (DHDPS) from *Mycobacterium tuberculosis.* Archives of Biochemistry and Biophysics 512(2): 154–159.

Fagen, J.R., Leonard, M.T., Coyle, J.F., McCullough, C.M., Davis-Richardson, A.G., Davis, M.J., Triplett, E.W. 2014. Liberibacter crescens gen. nov., sp. nov., the first cultured member of the genus Liberibacter. International Journal of Systematic and Evolutionary Microbiology 64: 2461-2466.

Fan, J., Chen, C., Yu, Q., Brlansky, R.H., Li, Z.G., and Gmitter, F.G. 2011. Comparative iTRAQ proteome and transcriptome analyses of sweet orange infected by '*Candidatus* Liberibacter asiaticus.' Physiologia Plantarum 143(3): 235-245.

Fares, M.A., Ruiz-González, M., Moya, A., Santiago, E.F., and Barrio, E. 2002. Endosymbiotic bacteria: *groEL* buffers against deleterious mutations. Nature 417:398.

Fares, M.A., Moya, A., and Barrio, E. 2004. GroEL and the maintenance of bacterial endosymbiosis. Trends in Genetics 20(9): 413-416.

Fink, A.L. 1998 Protein aggregation: folding aggregates, inclusion bodies and amyloid. Folding and Design 3(1): R9–R23.

Garcia-Olalla, C., and Garrido-Pertierra, A. 1987. Purification and kinetic properties of pyruvate kinase isoenzymes of *Salmonella typhimurium*. Biochemical Journal 241(2): 573–581.

Gilkes, J.M., Frampton, R.A., Smith, G.R., and Dobson, R.C.J. 2018 Potential pathogenicity determinants in the genome of '*Candidatus*' Liberibacter solanacearum, the causal agent of zebra chip disease of potato. Austalasian Plant Pathology 47(2): 119–134.

Goolsby, J.A., Adamczyk, J.J., Crosslin, J.M., Troxclair, N.N., Anciso, J.R., Bester, G.G., Bradshaw, J.D., Bynum, E.D., Carpio, L.A., Henne, D.C., Joshi, A., Munyaneza, J.E., Porter, P., Sloderbeck, P.E., Supak, J.R., Rush, C.M., Willett, F.J., Zechmann, B.J., and Zens, B.A. 2012. Seasonal population dynamics of the potato psyllid (Hemiptera: Triozidae) and its associated pathogen '*Candidatus* Liberibacter solanacearum' in potatoes in the southern great plains of North America. Journal of Economic Entomology 105(4): 1268-1276.

Griffin, M.D.W., Billakanti, J.M., Wason, A., Keller, S., Mertens, H.D.T., Atkinson, S.C., Dobson, R.C.J., Perugini, M.A., Gerrard, J.A., and Pearce, F.G. 2012. Characterisation of the first enzymes committed to lysine biosynthesis in *Arabidopsis thaliana*. PLoS One 7(7): e40318.

Hansen, A.K., Trumble, J.T., Stouthamer, R., and Paine, T.D. 2008 A new huanglongbing species, '*Candidatus* Liberibacter psyllaurous' found to infect tomato and potato, is vectored by the psyllid *Bactericera cockerelli* (Sulc). Applied and Environmental Microbiology 74: 5862–5865.

Hansen, J., and Cole, J. 1999. Analytical ultracentrifugation as a contemporary biomolecular research tool. Journal of Biomolecular Techniques 10(4): 163–176.

Hansen, J.C., Lebowitz, J., and Demeler, B. 1994. Analytical ultracentrifugation of complex macromolecular systems. Biochemistry 33(45): 13155–13163.

Hosokawa, T., Kikiuchi, Y., and Fukatsu, T. 2007. How many symbionts are provided by mothers, acquired by an offspring, and needed for successful vertical transmission in an obligate insect-bacterium mutualism. Molecular Ecology 16(24): 5316-5325.

Huang, C.Y., Lee, C.Y., Wu, H.C., Kuo, M.H., and Lai, C.Y. 2008. Interactions of chaperonin with a weakly active anthranilate synthase from the aphid endosymbiont *Buchnera aphidicola.* Microbial Ecology 56: 696–703.

Huot, O.B., Levy, J.G., and Tamborindeguy, C. 2018. Global gene regulation in tomato plant (Solanum lycopersicum) responding to vector (Bactericera cockerelli) feeding and pathogen ('*Candidatus* Liberibacter solanacearum') infection. Plant Molecular Biology 97(1-2): 57-72.

Ibanez, F., Levy, J., and Tamborindeguy, C. 2014. Transcriptome analysis of '*Candidatus* Liberibacter solanacearum' in its psyllid vector, *Bactericera cockerelli*. PLoS ONE 9(7): e100955.

Karsten, W.E. 1997. Dihydrodipicolinate synthase from *Escherichia coli*: pH dependent changes in the kinetic mechanism and kinetic mechanism of allosteric inhibition by L-lysine. Biochemistry 36(7): 1730–1739.

Lau, H.Y., Habibuddin, H., and Tan, C.S. 2011. Cloning, expression and purification of a partial outer membrane protein (Omp) from *Candidatus* Liberibacter asiaticus causing greening disease of citrus. Journal of Tropical Agriculture and Food Science 39(2): 179-189.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. 1992. Analytical ultracentrifugation in biochemistry and polymer science. The Royal Society of Chemistry 9: 0–125.

Lebowitz, J., Lewis, M.S., and Schuck P. 2009. Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Science 11(9): 2067–2079.

Leonard, M.T., Fagen, J.R., Davis-Richardson, A.G., Davis, M.J., and Triplett, E.W. 2012. Complete genome sequence of liberibacter crescens BT-1. Standards in Genomic Sciences 7: 7020271.

Levy, J.G., Mendoza, A., Miller, J.C., Tamborindeguy, C., and Pierson, E.A. 2017. Global gene expression in two potato cultivars in response to '*Candidatus* Liberibacter solanacearum' infection. BMC Genomics 18(1): 960.

Liefting, L.W., Perez-Egusquiza, C., and Clover, G.R.G. 2008b. A new '*Candidatus* Liberibacter' species in *Solanum tuberosum* in New Zealand. Plant Disease 92: 1474.

Liefting, L.W., Ward, L.I., Shiller, J.B., and Clover, G.R.G. 2008a A new *Candidatus* Liberibacter species in *Solanum betaceum* (tamarillo) and *Physalis peruviana* (cape gooseberry) in New Zealand. Plant Disease 92: 1588.

Lin, H., Lou, B., Glynn, J.M., Doddapaneni, H., Civerolo, E.L., Chen, C., Duan, Y., Zhou, L., and Vahling, C.M. 2011. The complete genome sequence of '*Candidatus* Liberibacter solanacearum', the bacterium associated with potato zebra chip disease. PLOS One 6: e19135.

Liu, H., Atta, S., and Hartung, J.S. 2017. Characterization and purification of proteins suitable for the production of antibodies against '*Ca*. L. asiaticus.' Protein Expression and Purification 139: 36–42.

Mafra, V., Martins, P.K., Francisco, C.S., Ribeiro-Alves, M., Freitas-Astua, J., and Machado, M.A. 2013. *Candidatus* Liberibacter americanus induces significant reprogramming of the transcriptome of the susceptible citrus genotype. BMC Genomics 14: 247.

Malcovati, M., and Valentini, G. 1982. AMP- and fructose 1,6-bisphosphate-activated pyruvate kinases from *Escherichia coli*. Methods in Enzymology 90: 170-179.

Mattevi, A., Valentini, G., Rizzi, M., Speranza, M.L., Bolognesi, M., and Coda, A. 1995. Crystal structure of *Escherichia coli* pyruvate kinase type I: molecular basis of the allosteric transition. Structure 3(7): 729–741.

Mazurek, S., Grimm, H., Boschek, C.B., Vaupel, P., and Eigenbrodt, E. 2002. Pyruvate kinase type M2: a crossroad in the tumor metabolome. British Journal of Nutrition 87(S1): S23–S29.

McCutcheon, J.P., and Moran, N.A. 2012. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10: 13–26.

Miles, G.P., Samuel, M.A., Chen, J., Civerolo, E.L., and Munyaneza, J.E. 2010. Evidence that cell death is associated with zebra chip disease in potato tubers. American Journal of Potato Research 87: 337–349.

Miran, A., and Moran, N.A. 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microbial Ecology 44: 137-143.

Moran, N.A. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108: 583–586.

Moran, N.A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proceedings of the National Academy of Sciences USA 93: 2873–2878.

Munyaneza, J.E. 2010. Psyllids as vectors of emerging bacterial diseases of annual crops. Southwestern Entomologist 35(3):471-477.

Munyaneza, J.E. 2012. Zebra chip disease of potato : biology, epidemiology and management. American Journal of Potato Journal 89:329–350.

Munyaneza, J.E., Goolsby, J.A., Crosslin, J.M.,and Upton, J.E. 2007. Further evidence that zebra chip potato disease in the Lower Rio Grande Valley of Texas is associated with *Bactericera cockerelli*. Subtropical Plant Science 59: 30–37.

Nachappa, P., Levy, J., and Tamborindeguy, C. 2012. Transcriptome analyses of *Bactericera cockerelli* adults in response to '*Candidatus* Liberibacter solanacearum' infection. Molecular Genetics and Genomics 287(10): 803-817.

Nwugo, C.C., Sengoda, V.G., Tian, L., and Lin, H. 2017. Characterization of physiological and molecular processes associated with potato response to Zebra chip disease. Horticulture Research 4: 17069.

Ochman, H., and Kelkar, Y.D. 2013. Genome reduction promotes increase in protein functional complexity in bacteria. Genetics 193: 303–307.

P.H., B., and Schuck, P. 2006. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. Biophysical Journal 90: 4651–4661.

Pagliai, F.A., Coyle, J.F., Kapoor, S., Gonzalez, C.F., and Lorca, G.L. 2017. LdtR is a master regulator of gene expression in *Liberibacter asiaticus*. Microbial Biotechnology 10(4):896-909.

Pagliai, F.A., Gardner, C.L., Bojilova, L., Sarnegrim, A., Tamayo, C., Potts, A.H., Teplitski, M., Folimonova, S.Y., Gonzalez, C.F., and Lorca, G.L. 2014. The transcriptional activator LdtR from *Candidatus* Liberibacter asiaticus mediates osmotic stress tolerance. PLoS Pathogens 10(4): e1004101.

Pagliai, F.A., Pan, L., Silva, D., Gonzalez, C.F., and Lorca, G.L. 2018. Zinc is an inhibitor of the LdtR transcriptional activator. PLoS One 13(4):e 0195746.

Peng, F., Widmann, S., Wunsche, A., Duan, K., Donovan, K.A., Dobson, R.C.J, and Lenski, R.E. 2018. Effects of beneficial mutations in *pykF* gene vary over time and across replicate populations in a long-term experiment with bacteria. Molecular Biology and Evolution 35(1): 202–210.

Ravindran, A., Saenkham, P., Gad-Levy, J., Tamborindeguy, C., Lin, H., Gross, D., and Pierson, E.A. 2017. Characterization of the serralysin-like gene of *Candidatus* Liberibacter

solanacearum associated with potato zebra chip disease. Phytopathology 108(3): 327–335.

Schuck, P. 2000. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modelling. Biophysical Journal 78:1606-1619.

Secor, G.A., Rivera, V.V., Abad, J.A., Lee, I., Clover, G.R.G., Liefting, L.W., Li, X., and De Boer, S.H. 2009. New diseases and epidemics association of '*Candidatus* Liberibacter solanacearum' with zebra chip disease of potato established by graft and psyllid transmission, electron microscopy and PCR. New Diseases and Epidemics 93: 330–335.

Secor, G.A., and Rivera-Varas, V.V. 2004. Emerging diseases of cultivated potato and their impact on Latin America. Rev Latinoa de La Papa 1: 1–8.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature 407: 81–86.

Slisz, A.M., Breksa, A.P., Mishchuk, D.O., McCollum, G., and Slupsky, C.M. 2012. Metabolomic analysis of citrus infection by '*Candidatus* Liberibacter' reveals insight into pathogenicity. Journal of Proteome Research 11(8): 4223-4230.

Sloan, D.B., and Moran, N.A. 2012 Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Molecular Biology and Evolution 29(12): 3781–3792.

Soares da Costa, T.P., Muscroft-Taylor, A.C., Dobson, R.C.J, Devenish, S.R.A., Jameson, G.B., and Gerrard, J.A. 2010. How essential is the 'essential' active-site lysine in dihydrodipicolinate synthase? Biochimie 92(7): 837–845.

Sreerama, N., and Woody, R.W. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Analytical Biochemistry 287(2): 252-260.

Tsumoto, K., Ejima, D., Kumagai, I., and Arakawa, T. 2003. Practical considerations in refolding proteins from inclusion bodies. Protein Expression and Purification 28: 1–8.

Tulloch, L.B., Morgan, H.P., Hannaeert, V., Michels, P.A.M, Fothergill-Gilmore, L.A., and Walkinshaw, M.D. 2008. Sulphate removal induces major conformational change in *Leishmania Mexicana* pyruvate kinase in the crystalline state. Journal of Molecular Biology 383(3): 615–626.

Turner, J.J, Healy, J.P., Dobson, R.C.J., Gerrard, J.A., Hutton, C.A. 2005. Two new irreversible inhibitors of dihydrodipicolinate synthase: diethyl (*E,E*)-4-oxo-2,5-heptadienedioate and diethyl (E)-4-oxo-2-heptenedioate. Bioorganic & Medicinal Chemistry Letters 15(4): 995–998.

van Ham, R.C.H.J., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J.M., Jimenez, L., Postigo, M., Silva, F.J., Tamames, J., Viguera, E., Latorre, A., Valencia, A., Moran, F., and Moya, A. 2003 Reductive genome evolution in *Buchnera aphidicola*. Proceedings of the National Academy of Sciences of the United States of America 100: 581–586.

Waygood, E.B., and Sanwal, B.D. 1974. The control of pyruvate kinases of *Escherichia coli* physicochemical and regulatory properties of the enzyme activated by fructose 1,6-diphosphate. Journal of Biological Chemistry 249: 265–274.

Wernegreen, J.J. 2002. Genome evolution in bacterial ensosymbionts of insects. Nat. Rev. Genet 3:850-861.

Wernegreen, J.J. 2015. Ensdosymbiont evolution: predictions from theory and surprises from genomes. Annals of the New York Academy of Sciences 1360(1):16–35.

Wernegreen, J.J., and Moran, N.A. 1999. Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein coding genes. Molecular Biology and Evolution 16(1): 83–97.

Whitmore, L., and Wallace, B.A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acid Research 32:668-73.

Whitmore, L., and Wallace, B.A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89(5): 392-400.

Workneh, F., and Paetzold, L. 2016. Population dynamics of released potato psyllids and their bacteriliferous status in relation to zebra chip incidence in caged field plots. Plant Disease 100(8): 1726-1767.

Wulff, N.A., Zhang, S., Setubal, J.C., Almeida, N.F., Martins, E.C., Harakava, R., Kumar, D., Rangel, L.T., Foissac, X., Bove, J.M., and Gabriel, D.W. 2014. The complete genome sequence of '*Candidatus* Liberibacter americanus' associated with citrus huanglongbing. Molecular Plant-Microbe Interactions 27(2): 163-176.

Yao, J., Saenkham, P., Levy, J., Ibanez, F., Noroy, C., Mendoza, A., Huot, O., Meyer, D.F., and Tamborindeguy, C. 2016. Interactions '*Candidatus* Liberibacter solanacearum'-Bactericera cockerelli: Haplotype effect on vector fitness and gene expression analyses. Frontiers in Cellular and Infection Microbiology 6: 62.

Yugari, Y., and Gilvarg, C. 1965. The condensation step in diaminopimelate synthesis. Journal of Biological Chemistry. 240: 4710.

Zhu, T., Bailey, M.F., Angley, L.M., Cooper, TF., and Dobson, R.C.J. 2010. The quaternary structure of pyruvate kinase type 1 from *Escherichia coli* at low nanomolar concentrations. Biochimie 92(1):116-120.

Zhu, T. 2008. Evolution of pyruvate kinase in the long-term evolution experiment of *Escherichia coli*: a structure/function study. PhD Thesis, University of Canterbury.

# Chapter Five

# Structural investigation of pyruvate kinase from *Ca*. L. solanacearum

The work presented in this chapter was undertaken as a collaboration between myself and an undergraduate student, Flynn Adcock, whom I supervised. Together, we screened and optimised crystallisation conditions of *Ca*. L. solanacearum pyruvate kinase. I solved the structure, and helped Flynn Adcock refine the structure.

## Preface

A major challenge for understanding *Ca*. L. solanacearum (*C*Lso) biology is the unculturable nature of the bacterium and the difficulty of isolating *C*Lso proteins *in vitro*. Chapter four described the development of a method to isolate soluble and functional recombinant *C*Lso proteins. This opens the door to a better understanding of CLso protein biology and allows us to characterise future drug targets for species specific antimicrobial design. Here, I introduce glucose metabolism in *C*Lso and discuss how the organism circumvents the first steps of glycolysis, Since the glucose-6-phosphoisomerase gene is absent from the *C*Lso genome. With *C*Lso *pykF* showing increased expression when *in planta* compared to in psyllids (Chapter Two), this Chapter highlights the importance of pyruvate kinase in glycolysis and discusses the potential of the enzyme as a novel antimicrobial target.

## 5.1 Introduction

### 5.1.1 Glucose metabolism in *Ca*. L. solanacearum

*C*Lso colonises the phloem tissue of tomato and potato plants, which is rich in sucrose, fructose and glucose (Karley et al. 2002; Viola et al. 2001). However, *C*Lso lacks a phosphotransferase system, which is required by many bacteria to transport fructose and sucrose across the inner membrane (Lin et al. 2011). However, the organism does have a glucose/galactose transporter, suggesting that glucose may be the major carbon source utilised by the bacterium (Lin et al. 2011). The *C*Lso genome also encodes a DctA-family dicarboxylate transporter that would allow for the uptake of succinate, fumarate, oxaloacetate and malate. Lin et al. (2011) propose that these compounds, malate in particular, may be utilised in potato plant colonisation in addition to glucose.

Although *C*Lso encodes most of the glycolytic pathway enzymes, a notable exception is the lack of glucose-6-phosphoisomerase (Lin et al. 2011). However, *C*Lso encodes a partial pentose phosphate pathway, which would allow the early glucose conversion steps of glycolysis to be bypassed. This would produce glyceraldehyde 3-phopshate in the oxidative phase of the pentose phosphate pathway, which is then feed into glycolysis to ultimately produce pyruvate and ATP (Lin et al. 2011) (Figure 5.1). As mentioned in Chapter One, it has

been proposed that *C*Lso relies on β-D-glucose for glycolysis instead of α-D-glucose. β-D-glucose is the sugar required for callose and cellulose synthesis in plants. Callose deposition is a well-known plant defence mechanism and has been observed *in planta* after *C*Lso infection. *C*Lso may use β-D-glucose as a source of carbon for glycolysis in order to reduce the amount of free β-D-glucose that could be used as a substrate for a pattern triggered immunity (PTI) response for plant cell wall fortification (Wang et al. 2017b). The essentiality of the glycolytic pathway for carbon metabolism as well as the use of β-D-glucose for pattern triggered immunity highlight enzymes in the pathway as potential antimicrobial targets.

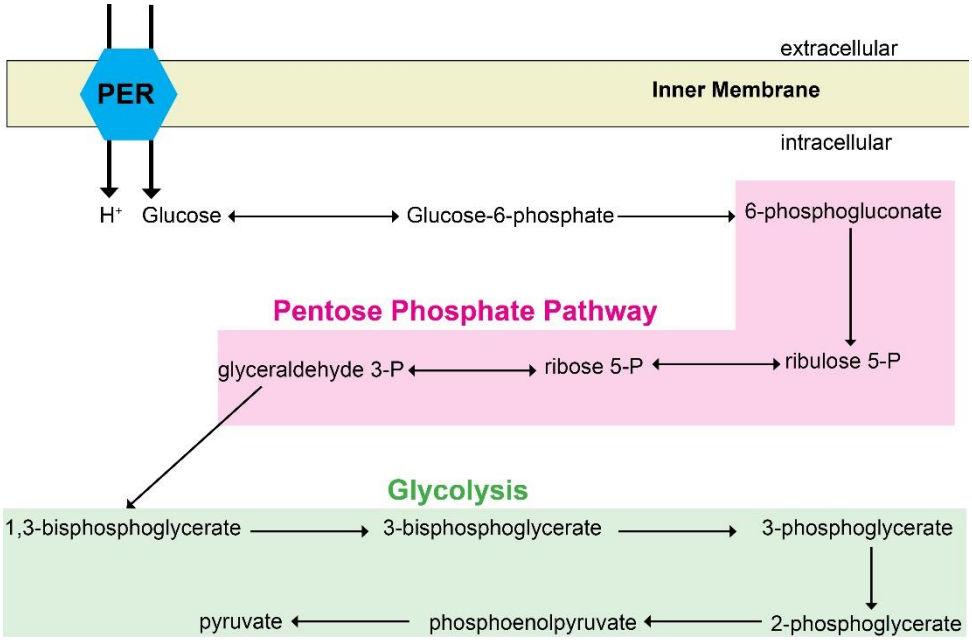



**Figure 5.1:** Schematic representation of glucose metabolism in *Ca*. L. solanacearum. Glucose is imported via glucose permease (PER) across the membrane. Glucose is converted to glucose-6-phosphate via glycolysis. Glucose-6-phosphate enters the pentose phosphate pathway as *C*Lso lacks glucose-6-phosphoisomerase.

Interestingly, the closest cultured relative of *C*Lso, *Liberibacter crescens*, has a considerably larger genome and retains a glucose-6-phosphoisomerase gene (Fagen et al. 2014). The absence of this gene in *C*Lso may indicate the richness of the intracellular environment that *C*Lso colonises leads to the loss of genes no longer required after the establishment of an endosymbiotic relationship with the psyllid.

### 5.1.2 Manipulation of psyllid energy metabolism

*C*Lso and *Ca*. L. asiaticus lack the key enzymes required at the terminal stages of oxidative phosphorylation including polyphosphate kinase and inorganic diphosphatase (Duan et al. 2009; Lin et al. 2011). Their absence, as well as the absence of terminal oxidases in the electron transport chain, attenuate the aerobic respiration capabilities for these bacteria (Duan et al. 2009; Lin et al. 2011). Interestingly, *C*Lso retains an ATP translocase nucleotide transport protein (*Ntta* gene) as well as ATP synthase, ATPase and ATP/ADP translocase genes (Lin et al. 2011). Thus, *C*Lso is capable of synthesising and scavenging ATP from the psyllid host (Lin et al. 2011; Killiny and Nehela 2017). However, despite the presence of an ATP translocase and synthase, Killiny and Nehela (2017) found that *Ca*. L. asiaticus increases ATP availability to the bacterium by inhibiting its breakdown in the insect host. An accumulation of ATP in *Ca*. L. asiaticus infected psyllids decreased the eukaryotes lifespan by inducing energetic stress on the vector, thus Killiny and Nehela (2017) hypothesised that the bacterium's pathogenicity could be due to metabolic imbalances caused by nutrient depletion and energy parasitism.

### 5.1.3 Pyruvate kinase

Glycolysis is the major pathway for the catabolism of monosaccharides in *C*Lso (Lin et al. 2011). Along with the absence of glucose-6-phosphoisomerase from the glycolytic pathway, *C*Lso lacks glucose 1-phosphatase and aldolase 1-epimerase, which are required for gluconeogenesis (Lin et al. 2011; Jain et al. 2017). However, the bacterium has retained a key glycolytic enzyme, pyruvate kinase (PK), which catalyses the final step in glycolysis via the irreversible conversion of phosphoenolpyruvate to pyruvate while also phosphorylating ADP to ATP (Valentini et al. 2000) (Figure 5.2). The reaction requires one monovalent cation ($K^+$) and two bivalent cations ($Mg^{2+}$) for catalytic efficiency (Mattevi et al. 1995). Pyruvate is a vital molecule as it acts as a branch point in a number of metabolic pathways. Under aerobic conditions, it is shuttled in the citric acid cycle or to lactic acid or alcoholic fermentation under anaerobic conditions (Garrett and Grisham 2012).

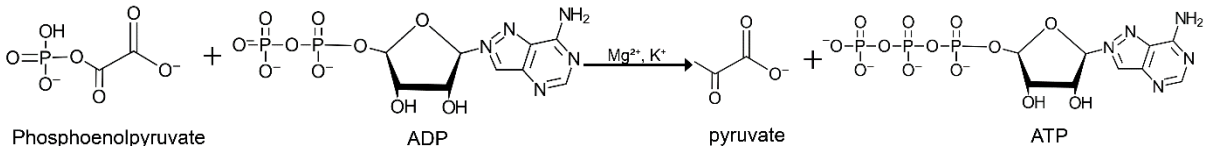



**Figure 5.2:** The reaction catalysed by pyruvate kinase.

Dufresne et al. (2005) found that genes that become non-essential are subject to relaxation of purifying selection and, over time, are unable to maintain function. With a functional ATP/ADP translocase and ATP synthase, it is unlikely that *C*Lso relies on glycolysis for production of ATP. Furthermore, Jain et al. (2017) hypothesised that pathogenic Liberibacters could utilise pyruvic acid as an entry into the citric acid cycle. In contrast, *C*Lso PK is functional (as discussed in Chapter Four) suggesting that there is a functional glycolytic pathway in *C*Lso.

5.1.3.1 Pyruvate kinase as a novel antimicrobial drug target

One strategy for antibacterial drug design has focused on targeting critical proteins and pathways unique to the bacterium. However, this significantly reduces the number of potential enzyme candidates (Kumar et al. 2012). Thus, alternative strategies focus on the inhibition of essential enzymes that catalyse the production of precursors to a variety of metabolic pathways (Zoraghi et al. 2011; Kumar et al. 2012). In the case of *C*Lso, genome reduction along with a lack of understanding of *C*Lso virulence, has severely limited the number of potential antimicrobial targets. The facultative intracellular pathogen *Brucella abortus* (cause of Brucellosis) has no classical virulence factors, such as cytolysins, exotoxins, capsules or fimbria (Seleem et al. 2008). However, Gao et al. (2016) found that PK was required for virulence in *B. abortus* as the enzyme allowed for intracellular survival and the establishment of chronic infection in mice, and for growth in nutrient deprived environments. In addition, PK is essential for *Staphylococcus aureus* viability and bacterial growth (Zoraghi et al. 2010). Because PK is essential for metabolism and growth, Zoraghi et al. (2011) propose the enzyme could serve as a novel drug target. Sequence divergence between *S. aureus* and eukaryotic PK genes has encouraged the possibility of a rational structure based approach to identify PK inhibitors (Zoraghi et al. 2011; Kumar et al. 2012).

Further, the bottleneck effect associated with organisms that adopt intracellular lifestyles results in the accumulation of nonsynonymous substitutions that increase the rate of

sequence evolution at both the nucleotide and amino acid level (McCutcheon and Moran 2012). Thus, sequence analysis does not easily reveal potential structural or catalytic changes of the enzyme in intracellular bacteria (McCutcheon and Moran 2012).

### 5.1.4 Overview of the chapter

The work in Chapter Two shows that pyruvate kinase is upregulated *in planta.* In addition, *Ca*. L. asiaticus has increased the gene expression of eight glycolysis related proteins *in planta*: glucokinase, fructose-bisphosphate aldolase, phosphoglycerate kinase, glyceraldehyde 3-phospahte dehydrogenase, phosphopyruvate hydratase, dihydrolipoamide dehydrogenase, phosphoenolpyruvate carboxykinase and phosphoglucomutase (Yan et al. 2013). Thus, glycolytic enzymes are thought to play an essential role in *C*Lso survival. In contrast, the absence of glucose-6-phosphoisomerase, the presence of ATP/ADP transporters and ATP synthase, as well as the low specific activity of *C*Lso PK demonstrated in Chapter Four raise the question as to whether these proteins have been affected by genetic drift and as a result have acquired nonsynonymous substitutions that alter PK catalytic efficiency or altering the structural integrity of the enzyme.

Here, I tested whether genetic drift had resulted in structural changes in *C*Lso PK that could account for the attenuated activity described in Chapter Four. This chapter presents, for the first time, a protein structure from *C*Lso. In addition, the Chapter further highlights the significance of chaperone protein co-expression in accelerating structural and functional understanding of *C*Lso proteins.

## 5.2 Results and Discussion

### 5.2.1 *C*Lso PK sequence is highly conserved among bacterial and eukaryotic homologues

A multiple sequence alignment was used to determine whether *C*Lso PK retained conserved residues at the catalytic and allosteric sites. The characteristic active site signature (Lys, Ile, Glu, Asn) is conserved in the *C*Lso gene, and the notable change of residue Asp249Lys found in *A. tumefaciens* is also changed *C*Lso (Figure 5.3). Similarly, there is also a change at Leu86 which is required for monovalent cation binding in *E. coli* to isoleucine in *C*Lso (Figure 5.3).

The consensus regions for PEP, ADP and the divalent cation are conserved between the bacterial species (Figure 5.3). Fructose-1,6-bisphosphate binds to the pyruvate kinase allosteric binding site in the C-domain which changes the conformation to allow enzyme activation (Waygood 1976; Jurica et al. 1998). The allosteric site of pyruvate kinase, although conserved among bacterial species, shows a change at position at Lys382 in *E. coli* to glutamine in *C*Lso (Gln383) (Figure 5.3). Interestingly, Lys382 is a key residue in fructose-1,6-bisphosphate binding (Mattevi et al. 1995; Donovan et al. 2016). This change in a residue considered important for fructose-1,6-bisphosphate binding raised the question of whether *C*Lso PK is allosterically regulated by the substrate.

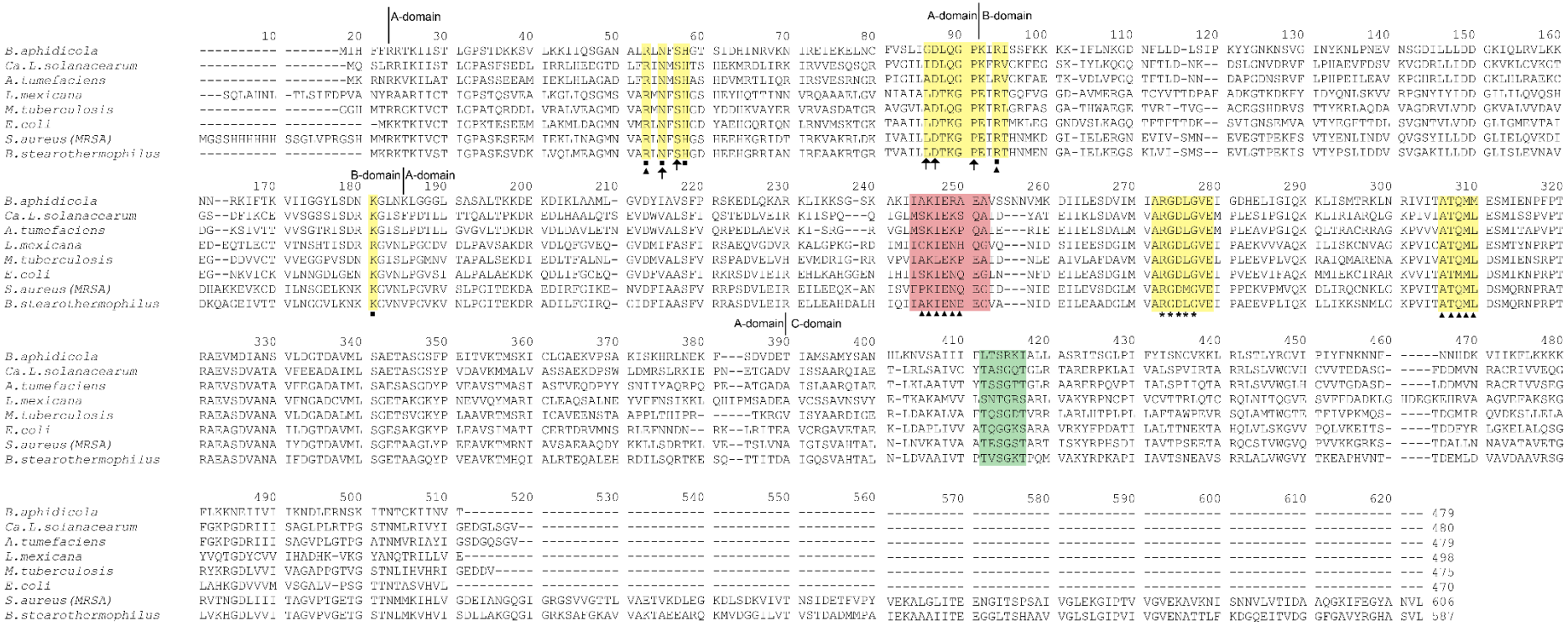



**Figure 5.3:** Sequence analysis of *C*Lso pyruvate kinase aligned with *Buchera aphidicola* (NP_240141.1), *Agrobacterium tumefaciens* (KIQ3038.1), *Leishmania mexicana* (Q27686.6), *Mycobacterium tuberculosis* (P9WKE5.1), *Escherichia coli* (P0AD61.1), methicillin resistant *Staphylococcus aureus* (CAG40767.1) and *Bacillus stearothermophilus* (CAA40994.1). Sites required for catalytic activity are yellow whereas key residues for the allosteric binding site are green. Sites required for binding ADP, PEP and monovalent cations are indicated by black squares, triangles and arrows respectively. Divalent cation binding sites are indicated by asterisks and are highlighted in red. Figured was generated in GENEIOUSr11.

### 5.2.2 Crystallisation of *C*Lso PK

*C*Lso PK expressed in Chapter Four was successfully crystallised using the MORPHEUS screen. Rectangular crystals were produced in condition B3 (90 mM sodium fluoride, sodium bromide, sodium iodide, 0.1 M imidazole and MES monohydrate (pH 6.5) and 30% v/v precipitant: 40% glycerol and 20% w/v PEG 4000) (Figure 5.4) (refer to Chapter Eight, Section 8.10). An X-ray diffraction dataset was collected to a maximum resolution of 3.0 Å. To select the high resolution cut-off for crystallographic refinement, the CC ½ criterion was used; this indicates the effective signal to noise of the data (Karplus and Diederichs 2015).

#### *5.2.2.1 Data processing and structure refinement*

Initial space group analysis suggested an orthorhombic ($P2_12_12_1$) space group. Data quality assessment using XTRIAGE from the PHENIX package (Adams et al. 2010) determined there was no evidence of twinning within the *C*Lso PK crystals (Figure 5.4). Molecular replacement was performed with the program PHASER (McCoy et al. 2007) from the CCP4 suite of programmes (Winn et al. 2011) using the monomer of *M. tuberculosis* PK as a search model (PDB ID: 5WRP) (Mattevi et al. 1995). Initial rounds of refinement were carried out in REFMAC5 (Murshudov et al. 1997) from the CCP4 program suite. Refinement of data in the $P2_12_12_1$ space group resulted in the lid domain of each subunit having unresolved electron density. The lid domain was removed and attempts were made to manually model the domain, but this also proved unsuccessful. Continued refinement provided $R_{free}$ values that would not decrease below 30% with 3% Ramachandran outliers (these were largely in the flexible lid domain, discussed later) and a clash score of 19.57. All relevant data collection and structure refinement statistics are provided in Table 5.1.

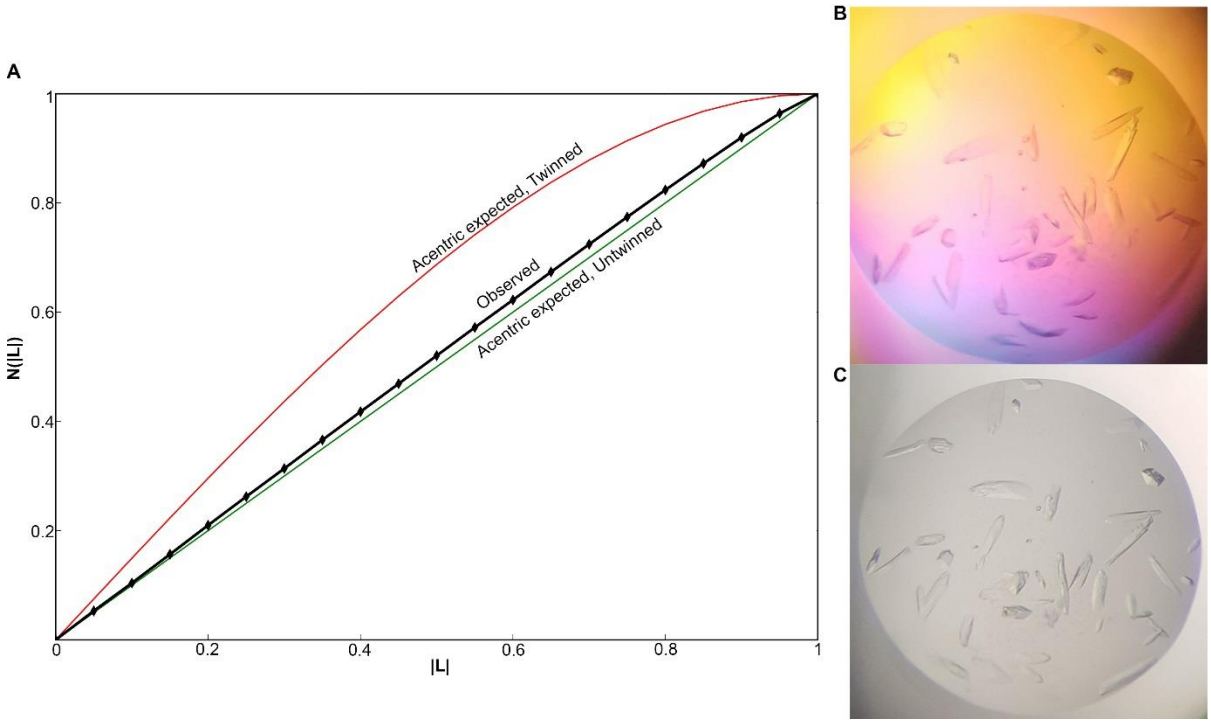



**Figure 5.4:** *L*-test for the detection of twinning in ligand free *C*Lso pyruvate kinase. **A** is $|L|$ plotted against $N(|L|)$. $L$ is given by $(I_1-I_2)/(I_1+I_2)$, where $I_1$ and $I_2$ are the intensities of unrelated reflections and $N(|L|)$ is the cumulative probability distribution of $L$. The *L*-test suggests the data is not twinned. **B** shows *C*Lso PK crystals under a light polarising filter to test for birefringence and **C** shows the variation in crystal morphologies.

**Table 5.1:** X-ray data collection and refinement statistics for *C*Lso PK. Statistical values for the highest resolution shells are given in parentheses. Each monomer within the asymmetric unit has a molecular weight of 55 kDa.

| **Data collection statistics** | **PK** |
|---|---|
| wavelength (Å) | 0.95369 |
| space group | $P2_12_12_1$ |
| unit cell parameters (Å, °) | a = 78.4, b = 164.69, c = 189.11, |
| resolution range (Å) | 49.13 – 3.08 (3.19 – 3.08) |
| observed reflections | 91715 (8711) |
| unique reflections | 45936 (4402) |
| mean *I/σ (I)* | 8.03 (1.12) |
| Completeness (%) | 0.97 |
| $R_{merge}$ | 0.056 (0.52) |
| $R_{meas}$ | 0.079 (0.73) |
| $R_{pim}$ | |
| CC ½ | 0.997 (0.68) |
| Wilson *B*factor (Å$^2$) | 84.16 |
| Molecules per asymmetric unit | 4 |
| **Structure and refinement statistics** | |
| $R_{factor}$ (%) | 0.258 (0.369) |
| $R_{free}$ (%) | 0.332 (0.389) |
| Number of atoms | |
| non-hydrogen | 11626 |
| macromolecules | 11626 |
| protein | 1519 |
| Solvent | 121 |
| Root-mean-square deviation (rmsd) | |
| bonds (Å) | 0.016 |
| angles (°) | 2.13 |
| Ramachandran favoured, allowed, outliers | 83, 14, 3 |
| Rotamer outliers | 13 |
| Clashscore | 19.57 |

### 5.2.3 The structure of *C*Lso PK

Consistent with the in-solution structure presented in Chapter Four, the crystal structure of *C*Lso PK is a homotetramer. The monomer is composed of three domains (A, B and C) (Figure 5.5). The A-domain has a characteristic $(\alpha/\beta)_8$-barrel fold with the active site located in a cleft between the A and B domain where the essential mono and divalent cations bind (Kumar and Barth 2011). The B-domain, also known as the lid domain lacked electron density resulting in poor modelling of residues in that region. Larsen et al. (1994) suggested that the B domain

rotates to close over the active site upon substrate binding. The C-domain is the allosteric domain and this consists of five α-helices and five strands of mixed β-sheets, creating the site where fructose-1,6-bisphosphate (allosteric activator) binds in order to activate the enzyme (Mattevi et al. 1995).

The *C*Lso PK tetrameric structure overlaid with homologues from *M. tuberculosis* (PDB ID: 5WRP) and *E. coli* (PDB ID: 1PKY) showed r.m.s.d. of 1.38 and 3.21 Å respectively indicating the presence of variations between the structures (Figure 5.5). The A- and C-domains are tightly packed, with the lid domain displaying different conformations between bacterial species (Figure 5.5).

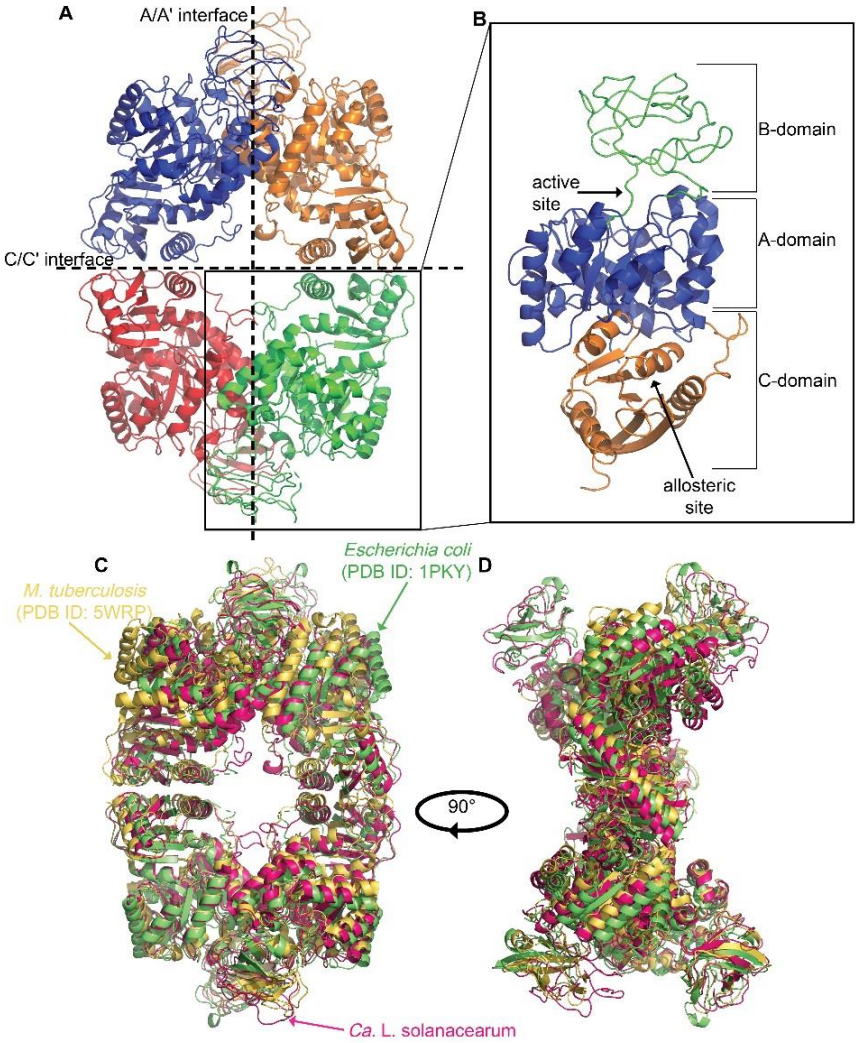



**Figure 5.5:** The structure of *C*Lso pyruvate kinase. **A** is the tetramer showing the orientation of the monomers (blue, orange, green and red). **B** is an enlargement of the monomer showing the location of the A, B and C domains as well as the active and allosteric sites. **C** and **D** are overlays of the *C*Lso (pink), *M. tuberculosis* (PDB ID: 5WRP) (yellow) and *E. coli* (PDB ID: 1PKY) (green) PK structures.

The crystal structure for *C*Lso PK is reported here at 3.0 Å. However, poor electron density failed to resolve residues within the lid domain of the protein. The backbone atoms of the PK lid domain are highly flexible (Donovan et al. 2016). A putty diagram revealed the relative atomic displacements of the tetramer showing the flexibility around the lid domain of *C*Lso PK (Figure 5.6). The high B-factors and resulting flexibility of the domain could account for the poor electron density observed (Huyton et al. 2003).

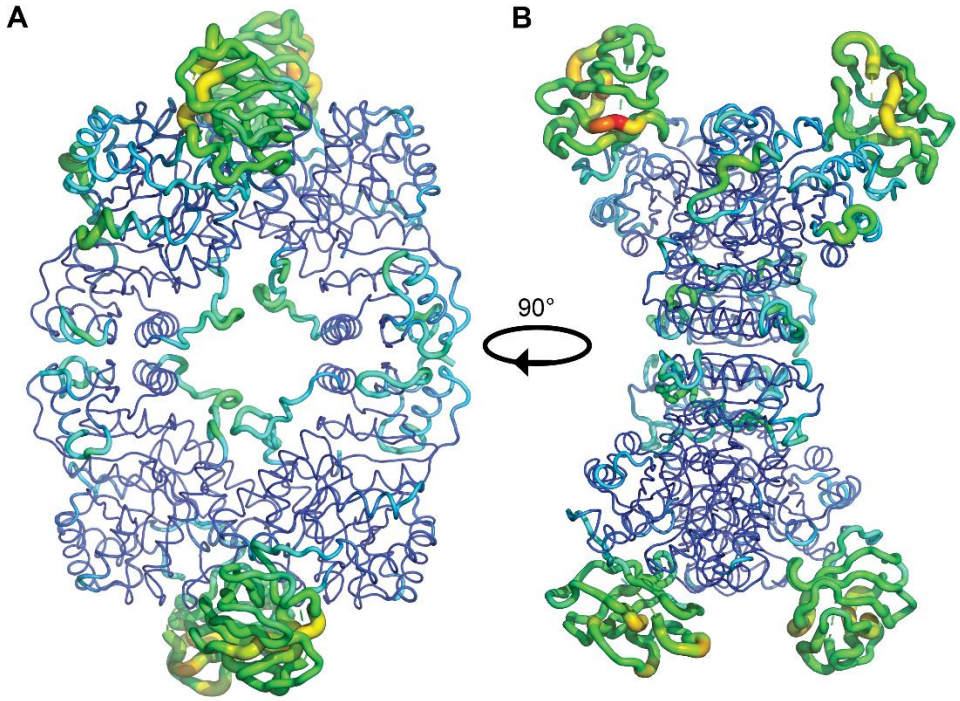



**Figure 5.6:** The tetrameric structure in the asymmetric unit of *C*Lso pyruvate kinase displayed by B-factor putty model to highlight relative atomic displacements (*Red*: high B-factor to *blue*: low B-factor).

Interface interactions are required for activation and subsequent conformational changes upon fructose-1,6-bisphosphate binding to PK (Jurica et al. 1998; Morgan et al. 2014). Analysis of the A/A' and the C/C' interfaces using PDBePISA (Krissinel and Henrick 2007) found the interface areas to be 1351 Å and 1272 Å, respectively, which is similar to the interface areas of *E. coli* PK of 1152 Å and 925 Å for A/A' and the C/C' interfaces, respectively (Donovan et al. 2016). The A/A' interface appears to be more stable, with 12 hydrogen bonds and 7 salt bridges across the interface compared to 8 hydrogen bonds and 2 salt bridges across the C/C' interface. Interestingly, the C/C' interface appears to be more hydrophobic, as the interface has a solvation energy of -3.5 kcal/mol compared to -18.2 kcal/mol for the A/A' interface (Fiorucci and Zacharias 2010).

*5.2.3.1 CLso PK retains residues required for catalytic activity*

To determine whether the attenuated catalytic activity reported in Chapter Four was a result of a change in residue orientation, which could affect substrate binding, the active site of the *C*Lso PK was analysed for structural differences. Consistent with the multiple sequence alignment, the key residues required for catalytic activity show very little change in orientation within the active site of *C*Lso PK. Interestingly, Thr278 required for PEP binding, has shifted approximately 3 Å away from the active site (Figure 5.7). This shift in residue position could explain the reduced PK activity of *C*Lso as the enzyme might not be able to bind PEP as tightly as other bacterial PK enzymes.

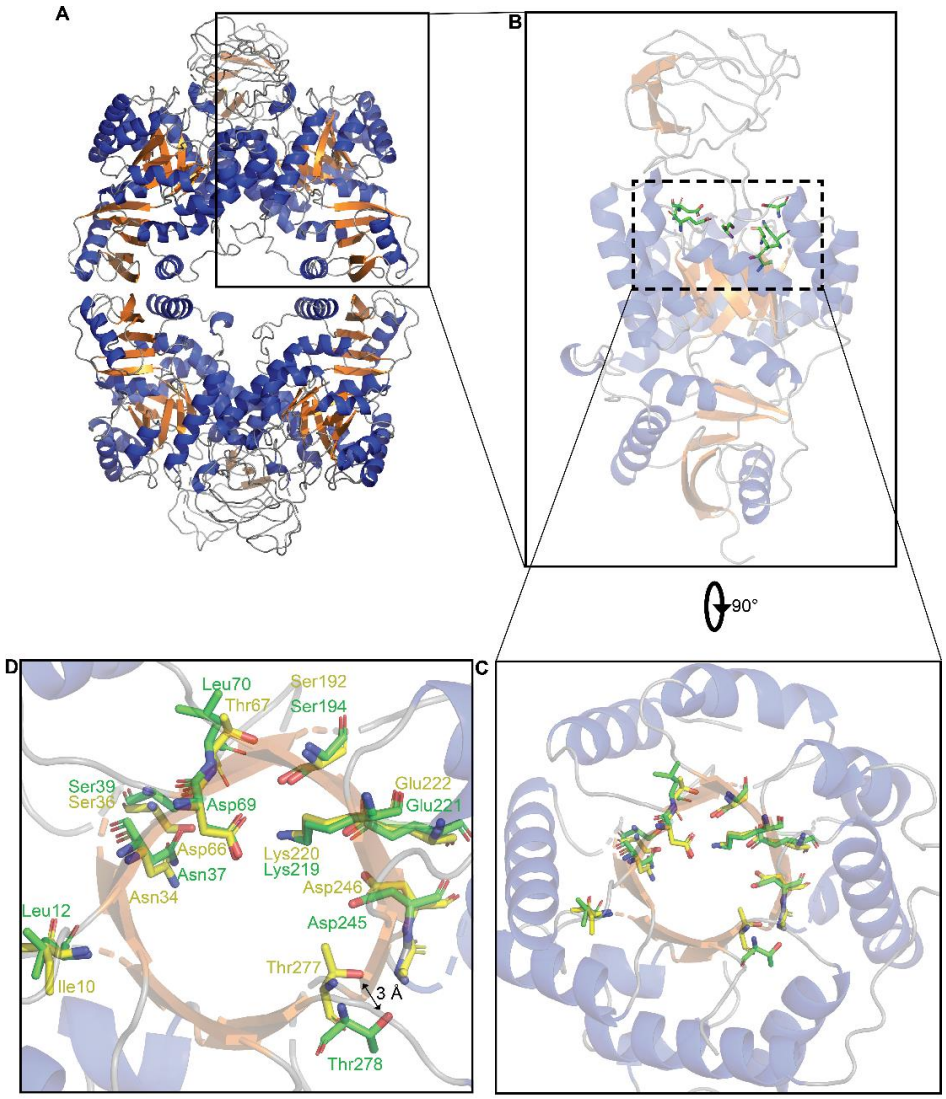



**Figure 5.7:** The active site of *C*Lso pyruvate kinase. **A** is the tetrameric structure and **B** is an enlargement of the monomer showing the residues required for catalytic activity. **C** and **D** are enlargements of the *C*Lso PK catalytic site (green) overlaid with *E. coli* PK (yellow, PDB ID: 1PKY).

*5.2.3.2 Residue changes in the allosteric site could account for reduced enzyme activity*

Pyruvate kinase is allosterically regulated, with binding of the allosteric activator inducing a conformational reorganisation of the PK tetramer from an inactive T-state to an active R-state (Morgan et al. 2014).

The fructose-1,6-bisphosphate allosteric binding site is located in a pocket formed from an effector loop in the C-domain. Ligand binding in *E. coli* is facilitated by residues Thr378, Gln379, Gly380, Gly381 and Lys382. Interestingly, the allosteric site for *C*Lso PK retains the conserved residues Thr379 and Gly382, but Ala380, Ser381 and Gln383 residues differ from

those in *E. coli* (Figure 5.8C). As Lys382 in *E. coli* is believed to be the key residue for fructose-1,6-bisphosphate binding, this raised the question as to whether the alternative Gln383 residue in *C*Lso PK could either prevent or alter the binding of the allosteric activator fructose-1,6-bisphosphate.

Here, the T-state is examined to gain an understanding of whether Gln383 in *C*Lso PK facilitated fructose-1,6-bisphosphate binding, the T-state *C*Lso PK structure was overlaid with *Saccharomyces cerevisiae* PK (PDB ID: 1AEW) with fructose-1,6-bisphosphate bound in the allosteric site (Figure 5.8D). Interestingly, Gln383 displayed polar contacts with the allosteric activator (Figure 5.8D).

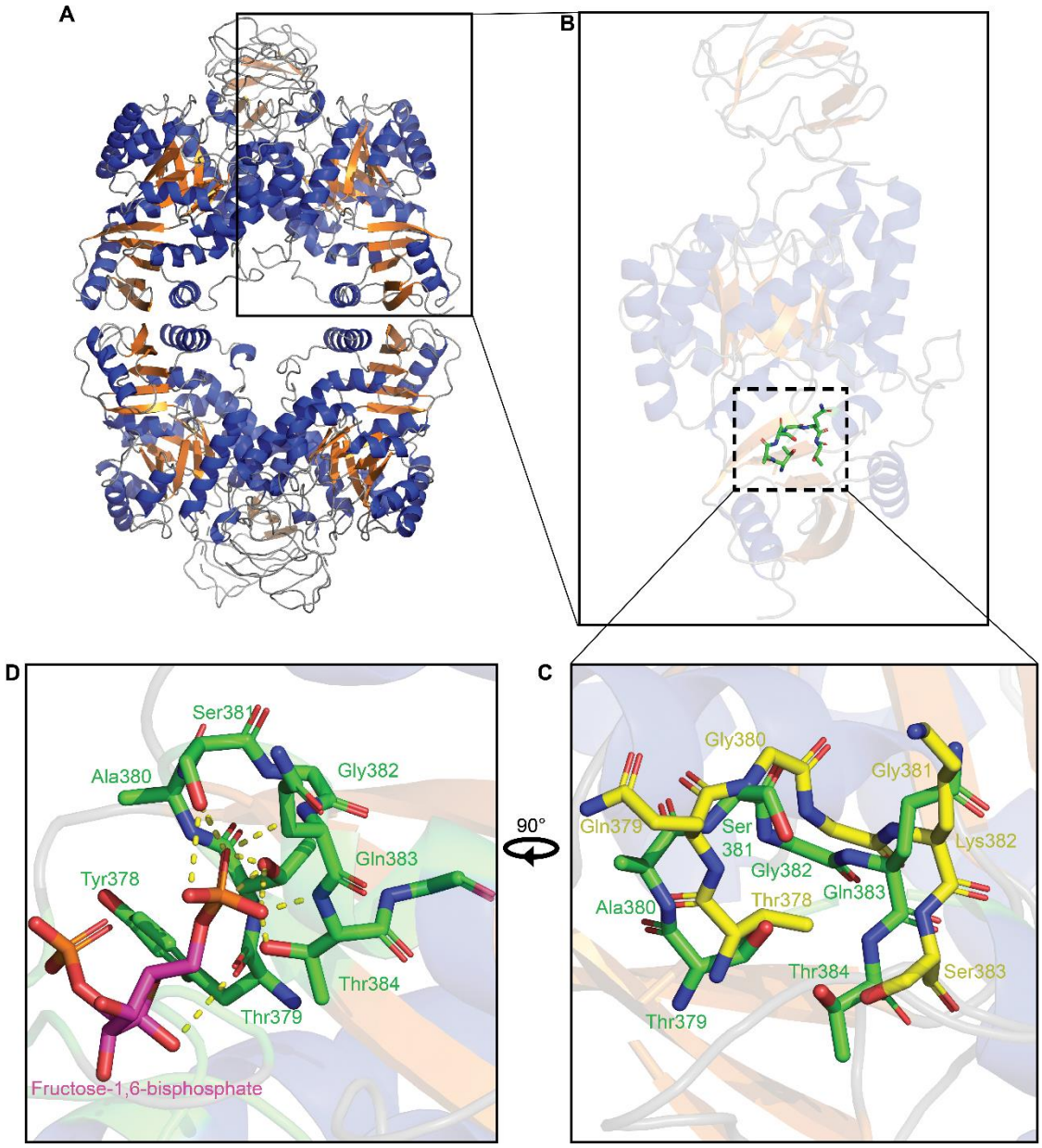



**Figure 5.8:** The allosteric site of *C*Lso pyruvate kinase. **A** is the tetrameric structure and **B** is an enlargement of the monomer showing the residues required for binding of fructose-1,6-bisphosphate in the allosteric site. **C** is an enlargement of the allosteric site overlaid with *E. coli* PK (yellow, PDB ID: 1PKY) and **D** shows the residues predicted to be involved in fructose-1,6-bisphosphate binding (overlaid with *S. cervesiae* PDB ID: 1A3W, with bound fructose-1,6-bisphosphate).

The multiple sequence alignment shows that the allosteric binding site of *C*Lso includes residue Gln383, in contrast to Lys382 in *E. coli*. Additionally, an overlay of the *C*Lso allosteric binding site shows that Gln383 is capable of binding fructose-1,6-bisphospahte. However, Valentini et al. (2000) deduce that in *E. col,i* when Lys382 is mutated to Gln, the glutamine mutant is unable to undergo a full transition to the R-state and that the lysine residue is

required for allosteric equilibrium and also that the substitution severely affects the functionality of the protein. Therefore, the *C*Lso Gln383 residue in the allosteric site might not be able to facilitate efficient fructose-1, 6-bisphosphate binding.

### *5.2.3.3* Small angle X-ray scattering *confirms the structure in solution and indicates a change in conformation when the allosteric activator binds*

Small angle X-ray scattering (*SAXS*) was used to verify that the crystal structure for *C*Lso PK is consistent with its solution structure and to determine whether fructose-1,6-bisphosphate induced a conformational change in structure (Figure 5.9A). Data quality was assessed by Guinier analysis of the SAXS scattering curves at low angles (Figure 5.9A inset) and revealed that no sample aggregation or inter-particle interference.

To determine whether the solution structure of *C*Lso PK correlated with the crystal structure, the experimental scatter for ligand-free *C*Lso PK, and the theoretical scatter generated from the X-ray crystal structure using CRYSOL were compared (Svergun et al. 1995). The scattering profile in solution agreed with the profile generated for the crystal structure, with a $\chi^2$ of 0.362 confirming that the crystal structure correlates with the solution structure for *C*Lso PK.

Insights from the *C*Lso PK crystal structure showed that the allosteric site was capable of binding fructose-1,6-bisphosphate. However, to test whether binding of the allosteric effector induces a conformational change in the protein, the experimental scatter for *C*Lso PK with fructose-1,6-bisphosphate was compared to the scatter for ligand-free PK. The $\chi^2$ for the fit was 0.26. Interestingly, the $D_{max}$ for *C*Lso PK with fructose-1,6-bisphosphate was 149.6, compared to the $D_{max}$ for ligand-free PK of 139.6 (previously reported in Chapter Four). This increase in $D_{max}$ could indicate a change in shape of the protein upon fructose-1,6-bisphosphate binding in the allosteric site (Figure 5.9B). The Kratky plot shows a bell shaped plot, which further indicates the protein is folded and suggests some flexibility in solution as the Kratky plot does not return to the baseline (Figure 5.9C). This is consistent with the mobil lid domains found in the structure.

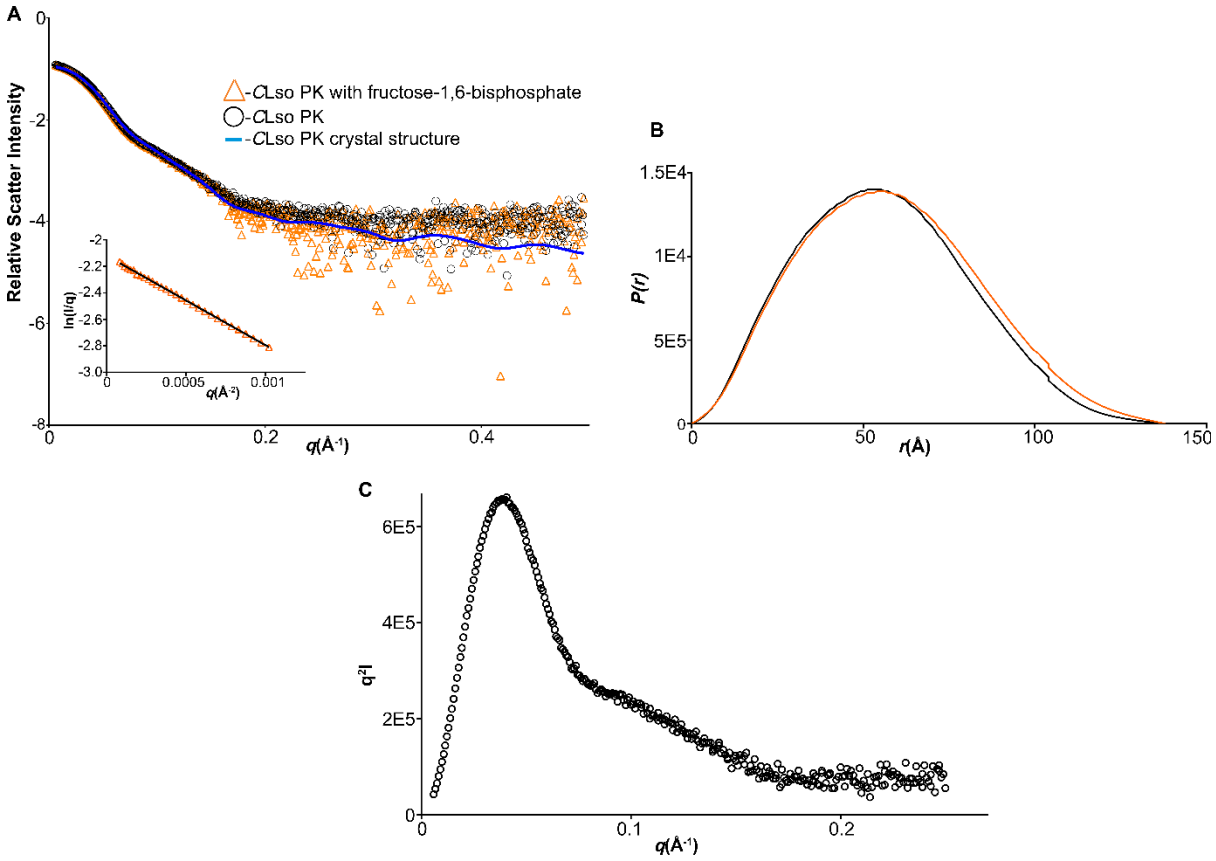



**Figure 5.9:** Small angle X-ray scattering data for *C*Lso pyruvate kinase. **A** is the experimental data presented as an intensity plot (*C*Lso PK ligand free represented as open circles, o) overlaid with the enzyme with fructose-1,6-bisphosphate (orange open triangles), and with the *C*Lso PK crystal structure CRYSOL fit to the data (in blue). The inset of **A** is the Guinier plot and **B** is the *P(r)* plot showing a change in structure of PK with fructose-1,6-bisphosphate in solution and **C** is the Kratky plot shows a bell shaped curve, which indicates the protein is folded.

## 5.3 Summary and future work

This chapter describes the structure of pyruvate kinase from *Ca.* L. solanacearum. The work identified a change in orientation of residue Thr278 in the active site compared to that of *E. coli.* This is hypothesised to affect phosphoenolpyruvate binding and could account for the low specific activity of the enzyme observed in Chapter Four. Future kinetic studies will aim to determine the kinetic parameters with respect to each of the substrates, phosphoenolpyruvate and ADP, in the presence and absence of fructose-1,6-bisphosphate to determine whether *C*Lso PK is allosterically activated. Structural binding analysis showed that Gln383 is situated at an optimal position for binding and plays a similar role to Lys382 in fructose-1,6-bisphospahte in the allosteric site. Future work will aim to mutate the *C*Lso

Gln383 to the Lys to determine whether the glutamine residue effects fructose-1,6-bisphosphate binding and activity of the enzyme.

I hypothesise that the presence of ATP synthase and ATP/ADP transporters, along with the loss of glucose-6-phosphoisomerase results in a relaxation of purifying selection upon enzymes in the *C*Lso glycolytic pathway. Further, genetic drift, known to occur in endosymbiotic bacteria, as well as an increase in GroEL and GroES expression in *C*Lso (Nachappa et al. 2014), result in the ability of *C*Lso PK to utilise a variety of substrates (Pollack et al. 2002; Donovan 2016). Future work should aim to kinetically characterise the enzyme with respect to various substrates in order to determine whether *C*Lso PK displays substrate versatility.

The chapter demonstrates that *C*Lso encodes a functional PK that is structurally homologous other bacterial homologues. Thus, the work presented here refutes the hypothesis of Jain et al. 2017 that proposed that *C*Lso encodes a non-functional glycolytic pathway. This further highlights the need for functional and structural investigation of *C*Lso enzymes, as inference of functional pathways cannot be solely made based on genome annotations that are unable to predict promiscuous enzymatic activities or alternative substrate specificity that result from genetic drift occurring in this intracellular bacterium.

## 5.4 References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., and McCoy, A.J. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66(2):213-221.

Donovan, K.A., Zhu, S., Liuni, P., Peng, F., Kessans, S.A., Wilson, D.J., and Dobson, R.C. 2016. Conformational dynamics and allostery in pyruvate kinase. Journal of Biological Chemistry 29*1*(17):9244-9256.

Duan, Y., Zhou, L., Hall, D.G., Li, W., Doddapaneni, H., Lin, H., Vahling, C.M., Gabriel, D.W., Williams, K.P., Dickerman, A., Shun, Y., and Gottwald, T. 2009. Complete genome sequence of citrus huanglongbing bacterium, '*Candidatus* Liberibacter asiaticus' obtained through metagenomics. Molecular Plant-Microbe Interactions 22(8): 1011-1020.

Dufresne, A., Garczarek, L., and Partensky, F. 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. Genome Biology 6(2):R14.

Fagen, J.R., Leonard, M.T., Coyle, J.F., McCullough, C.M., Davis-Richardson, A.G., Davis, M.J., and Triplett, E.W. 2014. *Liberibacter crescens* gen. nov., sp. nov., the first cultured member of the genus Liberibacter. International Journal of Systematic and Evolutionary Microbiology 64(7):2461-6.

Fiorucci, S., and Zacharias, M. 2010. Prediction of protein-protein interaction sites using electrostatic desolvation profiles. Biophysical Journal 98(9):1921-1930.

Gao, J., Tian, M., Bao, Y., Li, P., Liu, J., Ding, C., Wang, S., Li, T., and Yu, S. 2016. Pyruvate kinase is necessary for *Brucella abortus* full virulence in BALB/c mouse. Veterinary Research 47(1):87.

Huyton, T., Pye, V.E., Briggs, L.C., Flynn, T.C., Beuron, F., Kondo, H., Ma, J., Zhang, X., and Freemont, P.S. 2003. The crystal structure of murine p97/VCP at 3.6 Å. Journal of Structural Biology 144(3):337-348.

Jain, M., Munoz-Bodnar, A., and Gabriel, D.W. 2017. Concomitant loss of the glyoxalase system and glycolysis makes the uncultured pathogen "*Candidatus* Liberibacter asiaticus" an energy scavenger. Applied Environmental Microbiology 83(23):e01670-17.

Jurica, M. S., Mesecar, A., Heath, P. J., Shi, W., Nowak, T., and Stoddard, B. L. 1998. The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. Structure 6(2):195-210.

Karley, A.J., Douglas, A.E., *and* Parker, W.E. 2002. *Amino acid composition and nutritional quality of potato leaf phloem sap for aphids. Journal of Experimental Biology* 205**:***3009–3018.*

Karplus, P.A., and Diederichs, K. 2015. Assessing and maximizing data quality in macromolecular crystallography. Current Opinion in Structural Biology 34:60-68.

Killiny, N., and Nehela, Y. 2017. Metabolomic response to Huanglongbing: Role of carboxylic compounds in Citrus sinensis response to '*Candidatus* Liberibacter asiaticus' and its vector, Diaphorina citri. Molecular Plant-Microbe Interactions 30(8):666-678.

Krissinel, E., and Henrick, K. 2007. Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology 372(3):774-797.

Kumar, N.S., Amandoron, E.A., Cherkasov, A., Finlay, B.B., Gong, H., Jackson, L., Kaur, S., Lian, T., Moreau, A., Labrière, C., and Reiner, N.E. 2012. Optimization and structure–activity relationships of a series of potent inhibitors of methicillin-resistant *Staphylococcus aureus* (MRSA) pyruvate kinase as novel antimicrobial agents. Bioorganic & Medicinal Chemistry 20(24)7069-7082.

Kumar, S., and Barth, A. 2011. Effects of ions on ligand binding to pyruvate kinase: mapping the binding site with infrared spectroscopy. The Journal of Physical Chemistry B 115(20):6784-6789.

Larsen, T. M., Laughlin, L. T., Holden, H. M., Rayment, I., and Reed, G. H. 1994. Structure of rabbit muscle pyruvate kinase complexed with $Mn^{2+}$, $K^{+}$, and pyruvate. Biochemistry 33(20):6301-6309.

Lin, H., Han, C.S., Liu, B., Lou, B., Bai, X., Deng, C., Civerolo, E.L. and Gupta, G. 2013. Complete genome sequence of a Chinese strain of "*Candidatus* Liberibacter asiaticus". Genome Announcements 1(2):e00184-13.

Mattevi, A., Valentini, G., Rizzi, M., Speranza, M. L., Bolognesi, M. and Coda, A. 1995. Crystal structure of *Escherichia coli* pyruvate kinase type I: Molecular basis of the allosteric transition. Structure 3(7):729-741.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. 2007. Phaser crystallographic software. Journal of Applied Crystallography 40(4):658-674.

McCutcheon, J.P., and Moran, N.A. 2012. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10: 13–26.

Morgan, H. P., Zhong, W., McNae, I. W., Michels, P. A. M., Fothergill-Gilmore, L. A., and Walkinshaw, M. D. 2014. Structures of pyruvate kinases display evolutionarily divergent allosteric strategies. Royal Society of Open Science 1:140120.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D: Biological Crystallography 67(4):355-367.

Seleem, M.N., Boyle, S.M., and Sriranganathan, N. 2008. Brucella: a pathogen without classic virulence genes. Veterinary Microbiology 129(1)1-14.

Valentini, G., Chiarelli, L., Fortini, R., Speranza, M. L., Galizzi, A., and Mattevi, A. 2000. The allosteric regulation of pyruvate kinase: A site-directed mutagenesis study. Journal of Biological Chemistry 275(24):18145-18152.

Viola, R., Roberts, A.G., Haupt, S., Gazzani, S., Hancock, R.D, Marmiroli, N., Machray, G.C., and Oparka, K.J. 2001. Tuberization in potato involves a switch from apoplastic to symplastic phloem unloading. Plant Cell 13:385–398.

Waygood, E. B., Rayman, M. K., and Sanwal, B. 1975. The control of pyruvate kinases of *Escherichia coli*: Effectors and regulatory properties of the enzyme activated by ribose 5-phosphate. Canadian Journal of Biochemistry 53(4):444-454.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., and McNicholas, S.J. 2011. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D: Biological Crystallography 67(4):235-242.

Yan, Q., Sreedharan, A., Wei, S., Wang, J., Pelz-Stelinski, K., Folimonova, S. and Wang, N. 2013. Global gene expression changes in *Candidatus* L iberibacter asiaticus during the transmission in distinct hosts between plant and insect. Molecular Plant Pathology 14(4):391-404.

Zoraghi, R., See, R.H., Axerio-Cilies, P., Kumar, N.S., Gong, H., Moreau, A., Hsing, M., Kaur, S., Swayze, R.D., Worrall, L. and Amandoron, E. 2011. Identification of pyruvate kinase in methicillin-resistant *Staphylococcus aureus* as a novel antimicrobial drug target. Antimicrobial Agents and Chemotherapy 55(5):2042-2053.

# Chapter Six

## On the structure, function and regulation of dihydrodipicolinate synthase from *Ca*. L. solanacearum.

.

## Preface

This Chapter investigates the structure, function and allosteric regulation of *Ca*. L. solanacearum (*C*Lso) dihydrodipicolinate synthase. I introduce the relationship between the psyllid and the intracellular bacteria in maintaining genes for amino acid biosynthesis. Further, an introduction to lysine biosynthesis is discussed with a focus on the importance of dihydrodipicolinate synthase in the pathway, as it catalyses the committed step and is a potential antimicrobial target.

## 6.1 Introduction

### 6.1.1 Genome evolution in bacterial endosymbionts of insects

A transition to an endosymbiotic lifestyle within an insect results in a number of genomic evolutionary traits, most characteristic is the reduction in genome size and gene content of the intracellular bacteria (Andersson and Kurland 1998; Wernegreen 2002; McCutcheon and Moran 2012; Sloan et al. 2014). However, even with this genome reduction, many intracellular organisms retain genes required for the synthesis of essential amino acid biosynthesis (Sloan and Moran 2012). The endosymbiotic relationship has conferred the ability of eukaryotes to acquire metabolic capabilities from unique ecological environments (Wu et al. 2006). One such example is the obligate symbiont of aphids, *Buchnera aphidicola,* which allows the aphid to feed on sap from plant phloem (Wu et al. 2006). Phloem has few essential amino acids, which are unable to be synthesized by insects, and in turn *Buchnera* complements amino acid production for the aphids (Wernegreen 2002; Wu et al. 2006).

Despite these mutual benefits, the psyllid could also have subsequent alterations in fecundity, longevity and reproduction (Anderson and May 1982). *Ca*. L. asiaticus reduces the life span of adult *Diaphorina citri*, but fecundity is greater than uninfected psyllids (Pelz-Stelinski and Killiny 2016). Interestingly, Tamborindeguy et al. (2010) determined that *C*Lso-infected psyllids have greater longevity than unaffected counterparts, whereas *C*Lso-unaffected psyllid population have a greater population growth rate. Both *Ca*. L. asiaticus and *C*Lso have a close association between the bacterium and its vector, with transovarial and horizontal transmission occurring in both cases (Mira and Moran 2002; Pelz-Stelinski et al. 2010; Mann

et al. 2012; Workneh et al. 2016). Both Liberibacter species can colonise the majority of psyllid tissues, including reproductive organs (Ammar et al. 2011; Cicero et al. 2016). Overall, the effects of both *Ca*. L. asiaticus on *D. citri* and *C*Lso on *B. cockerelli* (psyllid) indicate a stable co-evolutionary association between the organisms (Pelz-Stelinski and Killiny 2016).

Psyllid feeding on potato tubers has shown to reduce the glutamic and aspartic acid levels within the plant, which are required for the synthesis of other amino acids such as lysine (Matthews and Hughes 1993; Bing-Yang et al. 2011). Bing Yang et al. (2011) determined that free amino acid composition was significantly reduced in potato plant tubers after feeding by *C*Lso-infected psyllids. *C*Lso possess relatively few enzymes required for *de novo* amino acids synthesis, which is complemented by the existence of three complete transport systems (*L*-amino acid ABC transporter system, proline/glycine-betaine ABC transporter system and a DctA-family dicarboxylate transporter) (Lin et al. 2011). However, *C*Lso does retain the genes for a complete lysine biosynthesis pathway (Lin et al. 2011).

Lysine is a member of the aspartate family of amino acids. Evolution has restricted the synthesis of aspartate-derived amino acids to fungi, bacteria and plants. Thus, animals must acquire lysine from their diet (Keeler et al. 1997; Viola 2001). Because lysine is an essential amino acid, inhibitors of lysine biosynthesis would be selectively toxic to plants and microorganisms, but not to animals. Thus, inhibitors of the lysine biosynthetic pathway may represent a new class of antimicrobial agents (Cox 1996; Coulter et al. 1999; Cox et al. 2000). The complexity of the *C*Lso and *B. cockerelli* endosymbiotic relationship is not fully understood, and it is not known if *C*Lso provides any selective advantage for the psyllid. Furthermore, *C*Lso is not the sole endosymbiont of *B. cockerelli* as *Candidatus* Carsonella ruddii, *Wolbachia*, *Acinetobacter* and *Methylibium* have all been identified as other bacterial endosymbionts of the psyllid (Nachappa et al. 2012), further complicating the understanding of the *C*Lso endosymbiotic relationship. Interestingly, these additional bacterial endosymbionts of *B. cockerelli* retain genes essential for lysine biosynthesis. Thus, an antibacterial agent that targets proteins of lysine biosynthesis could not only be specific for *C*Lso, but also for other bacterial endosymbionts of the psyllid and, depending on the endosymbiotic relationship and lysine requirement of *B. cockerelli*, a target specific for lysine biosynthesis could also indirectly effect the psyllid vector.

### 6.1.2 Lysine biosynthesis pathway

*C*Lso has lost many of its amino acid biosynthetic enzymes required for *de novo* synthesis of amino acids, however, the organism has retained the necessary genes for lysine biosynthesis, indicating the essentiality of this pathway for *C*Lso survival. Two pathways for the biosynthesis of lysine have been described; the α-aminoadipate pathway and the diaminopimelate pathway. The α-aminoadipate pathway is largely limited to higher order organisms such as fungi and euglenoids (Girodeau et al. 1986; Zabriskie and Jackson 2000). The pathway consists of the condensation of α-ketoglutarate and acetyl-CoA and proceeds to lysine via the (*S*)-α-aminoadipate intermediate (Zabriskie and Jackson 2000). Contrastingly, plants and bacteria synthesize lysine via the diaminiopimelate pathway (DAP) (Yugari 1962).

#### *6.1.2.1 The diaminopimelate pathway*

The condensation of pyruvate and (*S*)-aspartate-β-semialdehyde ((*S*)-ASA) to (4*S*)-hydroxy-2, 3, 4, 5-tetrahydro(2*S*)-dipicolinic acid (HTPA) characterises the first committed step of lysine biosynthesis in the diaminopimelate pathway (Blickling et al. 1997). It is then believed HTPA undergoes non enzymatic dehydration to dihydrodipicolinate (DHDP) (Blickling et al. 1997). The reaction is catalysed by dihydrodipicolinate synthase (DHDPS) and represents a critical point of regulation in the pathway as DHDPS is feedback inhibited by the end product of the reaction, lysine (Figure 6.1).

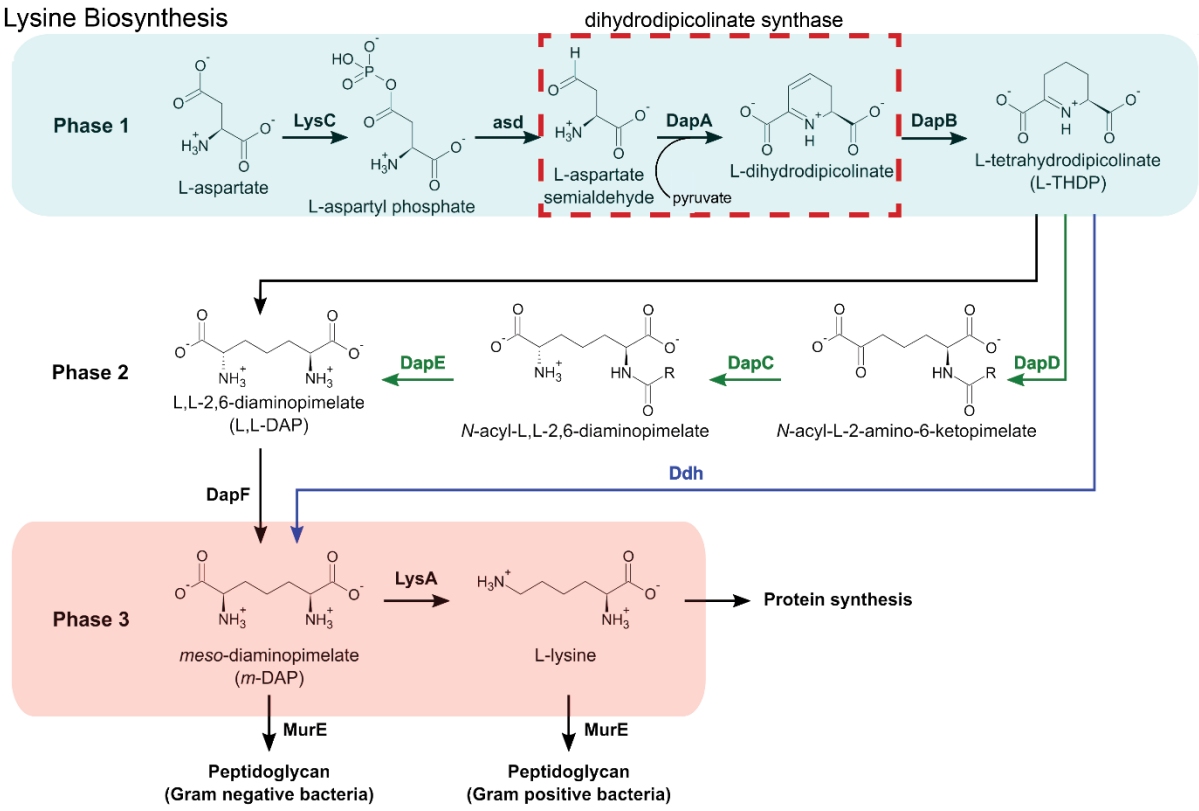



**Figure 6.1:** The three known variants of the diaminopimelate pathway

### 6.1.3 Dihydrodipicolinate synthase

Lysine and its precursor *meso*-DAP are major constituents of the bacterial peptidoglycan cell wall (Cox et al. 2000; Hutton et al. 2003). Lysine is a constituent of the cell wall in Gram-positive bacteria, whilst *meso*-DAP is required for Gram-negative bacteria (Mitsakos et al. 2008). Thus, a compound that inhibits the DAP pathway would be of a good inhibitor of bacterial growth (Coulter et al. 1999). DHDPS has received attention as a potential antibacterial target as it facilitates the committed step in lysine biosynthesis. Mitsakos et al. (2008) has reported progress in the development of species specific DHDPS inhibitors for *E. coli* and *M. tuberculosis* and methicillin resistant *Staphylococcus aureus* (MRSA). Interestingly, DHDPS is feedback inhibited by the final product of the DAP pathway, (*S*)-lysine with a number of studies determining that lysine acts allosterically on DHDPS (Yugari and Gilvarg 1965; Stahly 1969; Blickling et al. 1997).

### 6.1.4 Overview of the Chapter

*C*Lso persists as in two diverse but nutrient limited environments: as an endosymbiont within the psyllid and within the phloem of plants combined. Thus I propose that enzymes required for *C*Lso survival could be altered in order to adapt to a limited nutrient environment. The primary aim of this Chapter was characterise the catalytic mechanism of *C*Lso DHDPS. This Chapter details the kinetic characterisation of *C*Lso DHDPS with respect to its substrates: pyruvate and (*S*)-ASA and with the allosteric inhibitor, lysine. Kinetic studies show *C*Lso DHDPS has a high affinity for its substrates, and the potential benefits of this to the organism are discussed, providing insight into the relationship between *B. cockerelli* and *C*Lso with respect to lysine biosynthesis. Furthermore, this research presents the structure of *C*Lso DHDPS, the first structure of a protein from *C*Lso, with its substrates: pyruvate and succinic semi-aldehyde (an analogue of (*S*)-ASA) and with the allosteric inhibitor, lysine.

## 6.2 Results and Discussion

### 6.2.1 *Ca.* L. solanacearum DHDPS displays a ternary complex kinetic mechanism for substrate binding

No protein from the *C*Lso genome has been structurally or functionally characterised, thus, the organisms' biology and metabolism remains largely inferred from gene annotation. In order to design effective antibacterials it is imperative to have an understanding of the catalytic mechanism of a reaction. Therefore, in an attempt to understand the catalysis, regulation and substrate affinities of *C*Lso DHDPS, kinetic assays were undertaken.

A ping-pong kinetic generally describes the mechanism for DHDPS enzymes, based on the observation of initial velocity patterns that appear parallel (Kumpaisal et al. 1987; Karsten et al. 1997). Briefly, the ping-pong mechanism proceeds through a covalent enzyme-substrate complex with an irreversible step between the binding of the two substrates; this is usually the initial product (Cornish-Bowden 1999). In the case of DHDPS, the mechanism is proposed to involve the binding of pyruvate to free enzyme, followed by the release of a proton, which subsequently allows the binding of (*S*)-ASA to the enzyme and subsequent release of product and the regeneration of the free enzyme form (Karsten 1997) (Figure 6.2).

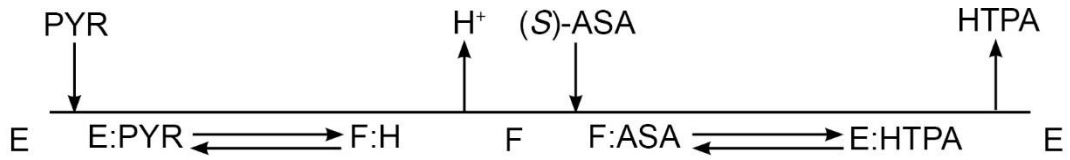



**Figure 6.2:** Schematic representation of the ping-pong mechanism of *Escherichia coli* DHDPS (Karsten 1997). Binding of pyruvate (PYR) to the free enzyme (E) leads to the formation of a Schiff base with the release of water. Subsequent tautomerisation of this Schiff base to an enamine results in the loss of a proton ($H^+$). (*S*)-ASA then binds to the stable substituted form (F) allowing product formation (HTPA) and the regeneration of the free enzyme.

In order to determine whether *C*Lso DHDPS follows a ping-pong kinetic model, initial rates with respect to varied (*S*)-ASA (0.002 - 0.128 mM) and pyruvate (0.002 – 0.08 mM) were measured using a coupled assay (described in detail in Chapter Eight, Section 8.9.2). Initial rates were taken in triplicate and the results were analysed using GraphPad Prism to determine the $K_m$ and $k_{cat}$ values. Figure 6.4 shows the kinetic plots, Lineweaver Burk plots (1/Rate versus 1/substrate concentration) and the standardised residual plots.

The initial rate data was first fitted to a ping-pong model (Equation 6.1), as this is the commonly observed mechanism known DHDPS homologues. The data was also fit to a ternary complex model (Equation 6.2) to determine whether there was an alternative binding mechanism.

$$v = VAB / (K_{mB}A + K_{mA}B + AB) \qquad \text{[Equation 6.1]}$$

$$v = VAB / (K_{mB}A + K_{mA}B + AB + K_{sA}K_{mB}) \qquad \text{[Equation 6.2]}$$

In equation 6.1 and 6.2, $V$ is the maximal velocity, $K_{mA}$ and $K_{mB}$ are the Michaelis-Menten constants for the two substrates, A and B are the respective substrate concentrations and $v$ is the initial velocity. Additionally, $K_{sA}$ is the binding constant for substrate A to the (E) enzyme form. In contrast to the ping-pong model, the ternary complex mechanism requires the enzyme to react with both substrates before the release of product (Figure 6.3).


**Figure 6.3:** Schematic representation of the ternary complex mechanism.

Data were fitted to appropriate models based on the fit of the non-linear regression and the lowest standard error ($R^2$) associated with the kinetic constants. Where the two models showed similar $R^2$ square values, an Akaike test was used to determine the model that best fit the kinetic data. The Akaike information criterion (AIC) is an estimate of the relative quality of a statistical model for a given data set relative to another model and provides a means for model selection (Akaike 1973). The corrected Akaike information criterion (AICc) is reported, which is the difference between the two corrected AIC values and is more accurate than AIC for smaller sample sizes (Akaike 1973).

Initial rate data for pyruvate and (*S*)-ASA fit the ternary complex model with an $R^2$ value of 0.99 (see Table 6.1). The Akaike test to determine the best fit kinetic model between ping pong and ternary complex determined that the probability the ping-pong model being correct was <0.01%, whereas the probability of the ternary complex model was 99.99% with a difference in the AICc value of 27.8. Previous kinetic analysis of DHDPS from *E. coli* (Kefala et al. 2008), *Staphylococcus aureus* (Girish et al. 2008), *Thermotoga maritima* (Pearce et al. 2006) and *Agrobacterium tumefaciens* (Atkinson et al. 2004) all fit initial rate data to a ping-pong model. In contrast, *C*Lso DHDPS displays a ternary complex mechanism, this is the first known occurrence of DHDPS demonstrating an alternative kinetic mechanism to the ping-pong model. The substrate inhibition model was also included in the analysis as DHDPS can be inhibited by (*S*)-ASA (Karsten 1997). Furthermore, Dobson (2003) determined that (*S*)-ASA prepared by ozonolysis (as used in this study) could cause substrate inhibition from the possible presence of artefacts associated with the substrate preparation (Karsten 1997).

The $K_m$ value for *C*Lso DHDPS with pyruvate is 0.012 ± 0.001 mM (Figure 6.4), whereas the $K_m$ with (*S*)-ASA is 0.025 ± 0.002 mM the $k_{cat}$ was 5.96 $s^{-1}$ (Table 6.1) and a $K_{sA}$ of 0.006. Interestingly, the kinetic constants generated from the regression were much lower for *C*Lso

when compared to those previously described for DHDPS in other organisms (Refer to table 6.2).

**Table 6.1**: Kinetic parameters for *C*Lso DHDPS with respect to pyruvate and (*S*)-ASA. The $K_m$, $V_{max}$ and $R^2$ for each fit were calculated in GraphPad Prism.

| Kinetic Parameters | Model | | |
|---|---|---|---|
| | **Ternary Complex** | **Ping-pong** | **Substrate inhibition** |
| $K_m$ (mM) | 0.012 | 0.016 | 0.068 |
| $V_{max}$ | 41.47 | 53.03 | 62.32 |
| $R^2$ | 0.99 | 0.98 | 0.95 |
| $k_{cat}$ ($s^{-1}$) | 5.96 ± 0.013 | | |
| $K_{mASA}$ (mM) | 0.025 ± 0.002 | | |
| $K_{mPyruvate}$ (mM) | 0.012 ± 0.001 | | |
| $k_{cat}$ / $K_{mASA}$ ($M^{-1}$ $s^{-1}$) | $2.4 \times 10^5$ | | |
| $k_{cat}$ / $K_{mPyruvate}$ ($M^{-1}$ $s^{-1}$) | $4.9 \times 10^5$ | | |
| $K_{sA}$ | 0.006 | | |

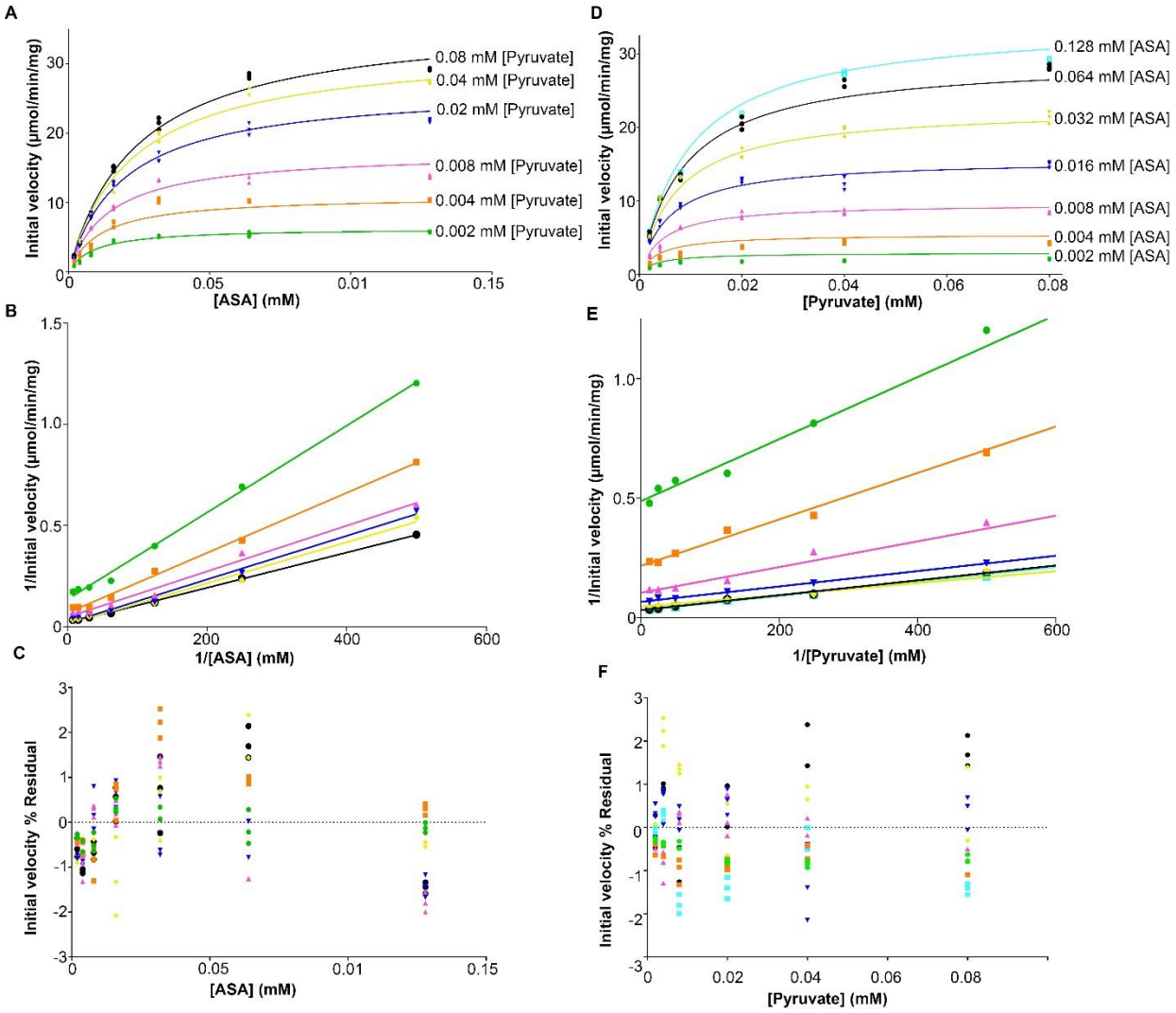



**Figure 6.4:** Kinetic plots of *C*Lso DHDPS. **A**, **B** and **C** are with respect to (*S*)-ASA at different pyruvate concentrations whereas **D**, **E** and **F** are with respect to pyruvate at different concentrations of (*S*)-ASA concentrations. **A** and **D** are direct plots, **B** and **E** are the Lineweaver-Burk plots, and **C** and **F** are the percentage residual plot of the data ($100 \times (v_{obs} - v_{cal})/v_{cal}$). All plots were generated in GraphPad Prism after non-linear regression of the data against the ternary complex model. The data has an $R^2$ value of 0.99.

**Table 6.2:** Comparison of kinetic constants between *Ca*. L. solanacearum DHDPS and other bacterial homologues

| Organism | $K_{m\ Pyruvate}$ (mM) | $K_{m\ ASA}$ (mM) |
|---|---|---|
| *Ca*. L. solanacearum | 0.012 | 0.025 |
| *Escherichia coli*[1] | 0.19 | 0.12 |
| *Bacillus subtilis* [2] | 1.07 | 3.13 |
| *Mycobacterium tuberculosis*[3] | 0.17 | 0.43 |
| *Staphylococcus aureus*[4] | 0.12 | 0.33 |
| *Thermotoga maritima*[5] | 0.053 | 0.16 |
| *Neisseria meningitides*[6] | 0.50 | 0.052 |
| *Agrobacterium tumefacien*[7] | 0.31 | 0.32 |
| *Arabidopsis thaliana*[8] | 1.0 | 0.09 |
| *Zea mays*[9] | 2.1 | 0.6 |

[1] – Karsten et al. 1997
[2] – Yamakura et al. 1974
[3] – Kefala et al. 2008
[4] – Girish et al. 2008
[5] – Pearce et al. 2006
[6] – Devenish et al. 2009
[7] – Atkinson et al. 2014
[8] – Griffin et al. 2012
[9] – Frisch et al. 1991

In the absence of a *B. cockerelli* (psyllid) genome, it is difficult to determine the endosymbiotic relationship between the psyllid and *C*Lso. Chapter Two identified an increase in gene expression of *C*Lso DHDPS in psyllid compared to plants, which could indicate that *C*Lso might provide lysine for the psyllid host and be essential for *C*Lso survival.

*C*Lso persists in two diverse environments: in the phloem of plants, an environment known to be low in the availability of amino acids (Wernegreen et al. 2002) and as an intracellular bacterium of the psyllid. The psyllid harbours additional bacterial endosymbionts such as *Candidatus* Carsonella ruddi, *Wolbachia* and *Acinetobacter* which could compete for nutrients and amino acids in the low nutrient environment of the psyllid. Harder and Dijkhuizen (1983) found that nutrient limited environments exert selective pressures that allow the organism to evolve mechanisms to survive within the restrictive environment. In order for the organism

to survive in the environment, it must be able to take up and metabolise the nutrient at the highest possible rate under conditions where the concentration outside the cell is low (Veldkamp and Jannasch 1972). One such mechanism proposed by Konings et al. (1980) is the ability for the organism to increase the rate of initial metabolism of the substrate that has accumulated inside the cell when the intracellular concentration is low; this could occur via the organism synthesising more of the enzyme involved in the initial metabolism of the substrate or alternatively, deriving a high affinity enzyme system. This correlates with the upregulation in psyllids of *dapA* encoding DHDPS compared to within the plant shown in Chapter Two; increase in gene expression indicate increased levels of DHDPS synthesis which could enhance the effectiveness of substrate utilisation for *C*Lso. Additonally, the low binding affinity of *C*Lso for its substrates could be an evolutionary mechanism evolved to respond to nutrient limitation in both the phloem and within the psyllid. Noell and Giovannoni (2019) described a similar occurrence for bacterioplankton which compete for resources in a nutrient depleted environment, and found that the bacteriums glycine-betaine transporter has an extraordinarily high affinity for glycine-betaine which is essential for metabolism and ATP production.

Endosymbiotic bacteria supplementing metabolites for insect hosts is a well-known occurrence (Wu et al. 2006). Insights into the catalysis of DHDPS in lysine biosynthesis demonstrate that lysine biosynthesis could be an important antibacterial target for *C*Lso, as the enzyme functions differently (*via* a ternary complex mechanism) and has significantly varying binding parameters compared to known DHDPS homologues. This suggests an alternative chemical mechanism for the DHDPS reaction compared to the ping-pong mechanism.

### 6.2.2 *C*Lso DHDPS is inhibited by lysine

DHDPS catalyses the committed step and represents a crucial point in the DAP pathway as it is feedback inhibited by the end product, lysine. Kinetic and structural studies have deduced that lysine is an allosteric inhibitor of DHDPS in *E. coli* (Karsten 1997). Although the precise mechanism by which lysine regulates DHDPS is still not clearly understood in bacteria. Interestingly, *Corynebacterium glutamicum* and *Bacillus anthracis* are insensitive to lysine inhibition (Rice et al. 2008; Domigan et al. 2009), which consequently raises the question of

how lysine biosynthesis is regulated in these organisms. In an attempt to determine whether *C*Lso DHDPS is feedback inhibited by lysine, initial rate data were collected at a range of inhibitor and substrate concentrations.

*6.2.2.1 Lysine displays mixed inhibition with respect to pyruvate*

Initial rate data were collected over a range of inhibitor and substrate concentrations. Lysine concentrations were varied from 0.05–0.8 mM and pyruvate concentrations ranged from 0.002–0.08 mM (see Chapter Eight, Section 8.9.2). The kinetic data were fitted to models that represent competitive, uncompetitive, non-competitive, mixed model, partial uncompetitive and partial non-competitive inhibition patterns to determine the best fit and the inhibition constant using GraphPad Prism (Table 6.3). A Lineweaver-Burk plot was used to show the trend of the data. The Akaike test for the best kinetic model of pyruvate with lysine between the mixed model inhibition model and partial non-competitive determined the probability that the respective models were correct were <99.99% and 0.01%, with a difference in the AICc values of 31.37. Inhibition of *C*Lso DHDPS by lysine using pyruvate as a substrate displayed mixed model inhibition with a $K_i$ of 0.38 ± 0.007 (Figure 6.5). The $K_m$ and $V_{max}$ for the mixed model inhibition were 0.012 ± 0.001 mM and 32.4 ± 0.48 µmol/min/mg, respectively.

**Table 6.3:** Kinetic parameters for lysine inhibition of *C*Lso DHDPS using pyruvate as a substrate. The $K_m$, $V_{max}$ and $R^2$ for each fit were calculated in GraphPad Prism.

| Kinetic Model | $K_m$ or $K_s$ (mM) | $K_i$ (mM) | $V_{max}$ (µmol/min/mg) | $V'$ (µmol/min/mg) | $R^2$ |
|---|---|---|---|---|---|
| mixed model | 0.012 ± 0.001 | 0.38 ± 0.007 | 32.4 ± 0.48 | | 0.99 |
| uncompetitive | 0.02 ± 0.001 | 0.32 ± 0.014 | 32.4 ± 0.83 | | 0.98 |
| non-competitive | 0.01 ± 0.003 | 0.15 ± 0.005 | 30.8 ± 0.43 | | 0.98 |
| uncompetitive | 0.013 ± 0.006 | 0.1 ± 0.004 | 33.8 ± 0.58 | | 0.98 |
| partial uncompetitive | 0.002 ± 0.005 | 0.3 ± 0.0012 | 33.8 ± 0.45 | | 0.97 |
| competitive | 0.006 ± 0.001 | 0.024 ± 0.004 | 27.1 ± 1.14 | | 0.87 |

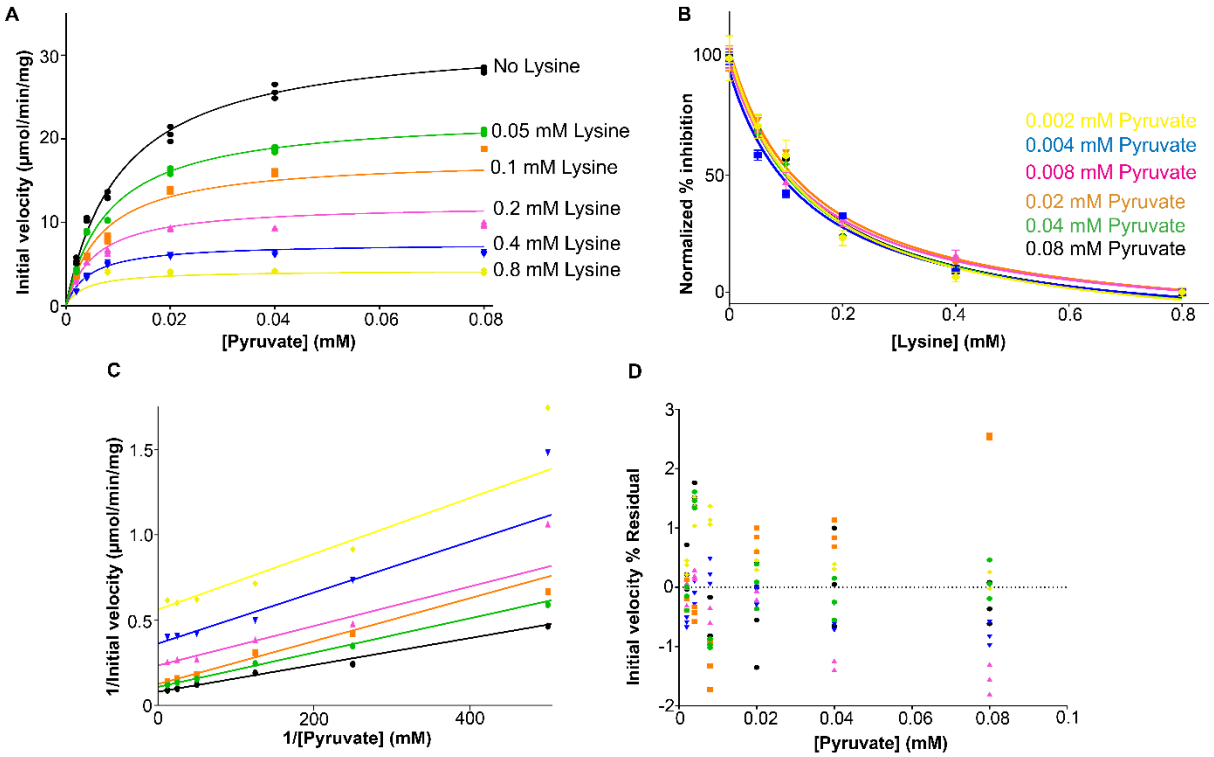



**Figure 6.5:** Kinetic plots of *C*Lso DHDPS. **A** is a direct plot with respect to pyruvate at different lysine concentrations fitted to a mixed model inhibition whereas **B** is the percentage inhibition at varying lysine concentrations compared to the rate in the absence of lysine. The error bars are the standard deviation. **C** is the Lineweaver-Burk plot and **D** is the percentage residual plot of the data (100 × ($v_{obs}$ – $v_{cal}$)/$v_{cal}$). GraphPad Prism generated all plots after non-linear regression of the data against the ternary model. The data has an $R^2$ value of 0.99.

Mixed inhibition occurs when the inhibitor binds to the enzyme whether or not the enzyme has already bound the substrate (Cornish Bowden 1999). Equation 6.3 is the steady state expression that describes the mixed model inhibition, which is illustrated in Figure 6.6.

$$v = V/((s + i/K'_i) + (1 + (K_s/s)) + (1 + (i/K_i)))$$ [Equation 6.3]

Where *s* and *i* represent the concentrations of substrate and inhibitor respectively, and $K_i$, $K'_i$, $K_s$ are dissociation constants.

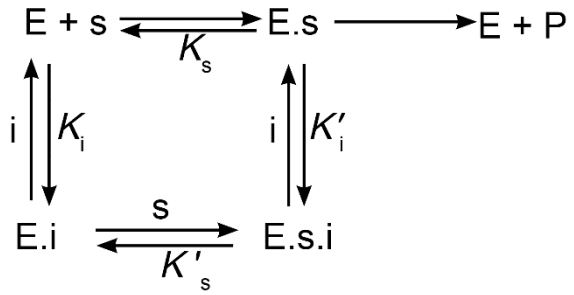



**Figure 6.6:** Schematic representation of mixed inhibition model where E represents the enzyme, s the substrate and i the inhibitor.

*6.2.2.2 Lysine shows uncompetitive inhibition with respect to (S)-ASA*

Initial rate data was collected over a range of inhibitor and substrate concentrations. Lysine concentrations were 0.05–0.8 mM and (*S*)-ASA concentrations ranged from 0.002–0.064 mM (see Chapter Eight, Section 8.9.2). The kinetic data were fitted to following models using GraphPad Prism; competitive, uncompetitive, non-competitive, mixed, partial uncompetitive and partial non-competitive inhibition (see Table 6.4). A Lineweaver-Burk plot was used to show the quality of the data (Figure 6.7). The Akaike test for the best kinetic model of (*S*)-ASA with lysine between the uncompetitive and mixed model inhibition determined the probability of the mixed model being correct was 28%, whereas the probability of the uncompetitive model being correct was 72% (AICc value of 1.8).

Inhibition of *C*Lso DHDPS by lysine with (*S*)-ASA as a substrate displayed uncompetitive inhibition with a $K_i$ of 0.32 (Figure 6.7). The $K_m$ and $V_{max}$ for the competitive inhibition model were 3.7 ± 0.1 mM and 37.8 ± 0.4 µmol/min/mg, respectively.

**Table 6.4:** Kinetic parameters for lysine inhibition of *C*Lso DHDPS with respect to (*S*)-ASA. The $K_m$, $V_{max}$ and $R^2$ for each fit were calculated in GraphPad Prism.

| Kinetic Model | $K_m$ or $K_s$ (mM) | $K_i$ (mM) | $V_{max}$ (µmol/min/mg) | $V'$ (µmol/min/mg) | $R^2$ |
|---|---|---|---|---|---|
| uncompetitive | 0.02 ± 0.001 | 0.32 ± 0.014 | 32.4 ± 0.83 | | 0.98 |
| mixed model | 0.02 ± 0.02 | 0.33 ± 0.03 | 30.37 ± 0.9 | | 0.97 |
| partial non-competitive | 0.008 ± 0.003 | 0.35 ± 0.02 | 25.9 ± 0.85 | 3.8 ± 0.74 | 0.97 |
| partial uncompetitive | 0.008 ± 0.001 | 0.58 ± 0.01 | 27.94 ± 0.85 | | 0.96 |
| non-competitive | 0.008 ± 0.001 | 0.77 ± 0.03 | 21.36 ± 0.84 | | 0.96 |
| competitive | 0.0125 ± 0.002 | 0.17 ± 0.03 | 26.09 ± 1.6 | | 0.87 |

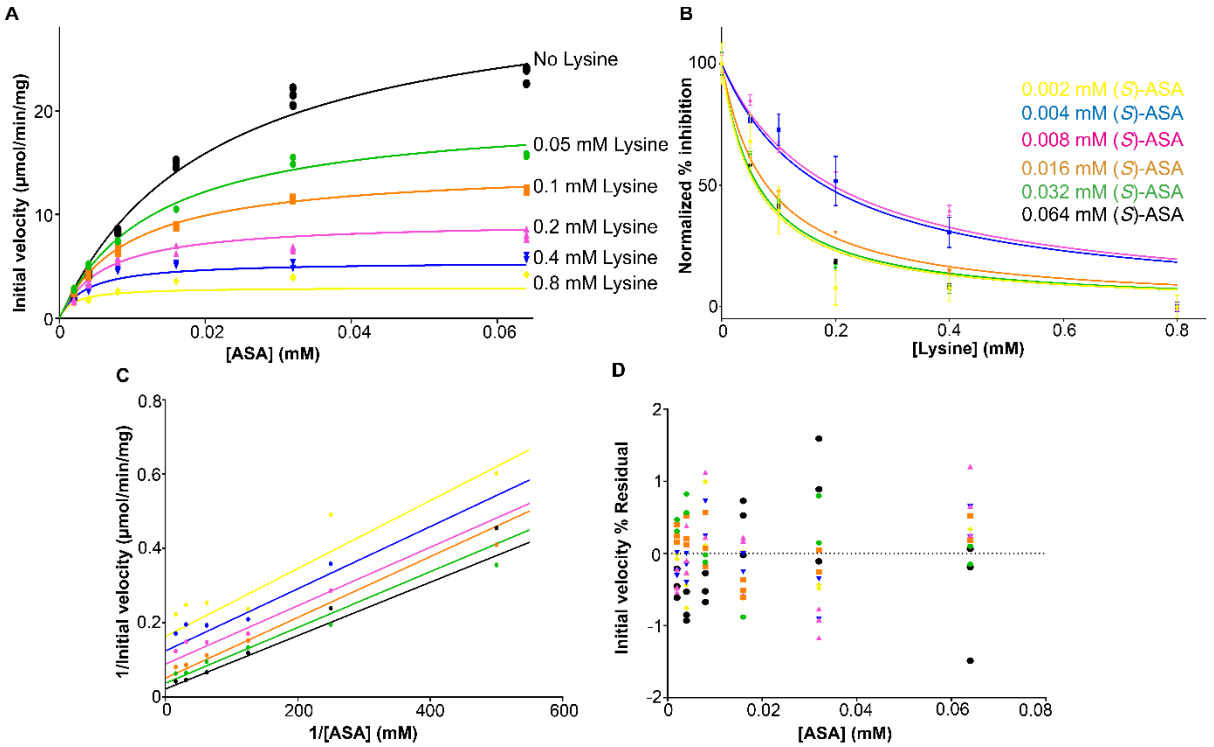



**Figure 6.7:** Kinetic plots of *C*Lso DHDPS **A** is with respect to (*S*)-ASA at different lysine concentrations fitted to a uncompetitive inhibition model whereas **B** is the percentage inhibition at varying lysine concentrations compared to the rate in the absence of lysine. The error bars are the standard deviation. **C** is the Lineweaver-Burk plot and **D** is the percentage residual plot of the data ($100 \times (v_{obs} - v_{cal})/v_{cal}$). GraphPad Prism generated all plots after non-linear regression of the data against the ternary model. The data has an $R^2$ value of 0.98.

Uncompetitive inhibition decreases the apparent value of V with no effect on V/$K_m$ (Cornish-Bowden 1999). Equation 6.4 is the steady state expression that describes an uncompetitive inhibitor, which is illustrated in Figure 6.8. Uncompetitive inhibition is predicted when the inhibitor binds only to the enzyme/substrate complex and not to the free enzyme.

$v = V/(1+(i/K_i)+(K_m/s))$ [Equation 6.4]

Where *s* and *i* represent the concentrations of substrate and inhibitor respectively, and $K_i$ is a dissociation constant.

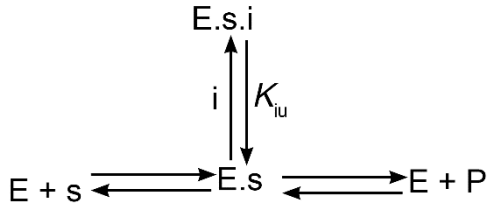



**Figure 6.8:** Schematic representation of uncompetitive inhibition where E represents the enzyme, s the substrate and i the inhibitor.

To date, many characterised DHDPS enzymes are known to be partially inhibited by lysine (Dobson et al. 2004; Atkinson et al. 2014). The Chapter demonstrates that lysine is not a partial inhibitor of *C*Lso DHDPS, which raised the question as to whether the binding of lysine in the allosteric site differs compared to other known DHDPS enzymes.

### 6.2.3 The crystal structure of *Ca*. L. solanacearum DHDPS

Kinetic characterisation of *C*Lso DHDPS determined that the enzyme had a higher affinity for (*S*)-ASA and pyruvate than DHDPS homologues. Between the biophysical characterisation presented in Chapter Four and the functional studies presented in this Chapter, there is no obvious explanation for this increased affinity for substrates. Thus, it was proposed that structural changes in binding of the substrates could account for this. To address this, X-ray crystal structures of *C*Lso DHDPS without substrates, with pyruvate and with pyruvate and succinic semi-aldehyde (SSA), (an analogue of (*S*)-ASA) were solved, in order to investigate the substrate affinity of the enzyme in the context of atomic resolution structural data.

Currently, there are no known crystal structures for *C*Lso proteins. Furthermore, structures of proteins in Liberibacter species are limited, hindering an understanding of this genus'

biochemistry and biology. A major obstacle to obtaining protein structures from Liberibacter is the difficulty in producing soluble protein. In this section, I present the first crystal structure of the *C*Lso protein, DHDPS, using the method previously described in Chapter Four (detailed method provided in Chapter Eight Sections 8.6.1.4 and 8.7.2).

*6.2.3.1 Crystallisation and diffraction data collection*

Protein crystals of *C*Lso DHDPS with and without ligands was successfully obtained using the JCSG+ Suite screen. Rectangular crystals were produced in condition A2 (20% (w/v) polyethylene glycol 3000, 0.1 M sodium citrate (pH 5.5)) (refer to Chapter Eight, Section 8.10). To select the high resolution cutoff for crystallographic refinement, the CC½ value was used which indicates the effective signal to noise of the data (Karplus and Diederichs 2015). X-ray diffraction datasets were collected at: 2.4, 2.01, 2 and 2.4 Å for *C*Lso DHDPS with pyruvate, *C*Lso DHDPS with lysine, with pyruvate and with succinic semi-aldehyde and pyruvate and for ligand free *C*Lso DHDPS, respectively.

*6.2.3.2 Data processing and structure refinement*

6.2.3.2.1 Obstacles in solving the structure of ligand free *C*Lso DHDPS

Data was processed in the spacegroup $C222_1$ and solved by molecular replacement using the monomer of *Agrobacterium tumefaciens* DHDPS as the search model (PDB ID: 4I7U, 60% sequence identity) (Atkinson et al. 2012). In this spacegroup, numerous rounds of refinement failed to reduce the residual factor ($R_{factor}$) below 37%. This was believed to be either because the data was twinned as a result of growth of the crystal in which two separate crystals share some of the crystal lattice points in a symmetrical manner (Figure 6.8) or in the wrong space group. To resolve the spacegroup, the structure was solved in the C2 space group, but this still failed to resolve the $R_{factor}$. Data was scaled in a lower symmetry monoclinic space group (P1). Here, the asymmetric unit contains eight monomers, corresponding to two tetramers. Attempts to refine the structure in this space group failed to resolve the $R_{factor}$. As a result, XTRIAGE (Zwart et al. 2005) from the PHENIX package (Adams et al. 2010) was used to determine whether the *C*Lso DHDPS in a ligand free crystal was twinned or processed in the

incorrect space group. The *L*-test did not indicate twinning for the *C*Lso ligand free DHDPS data (Figure 6.9).

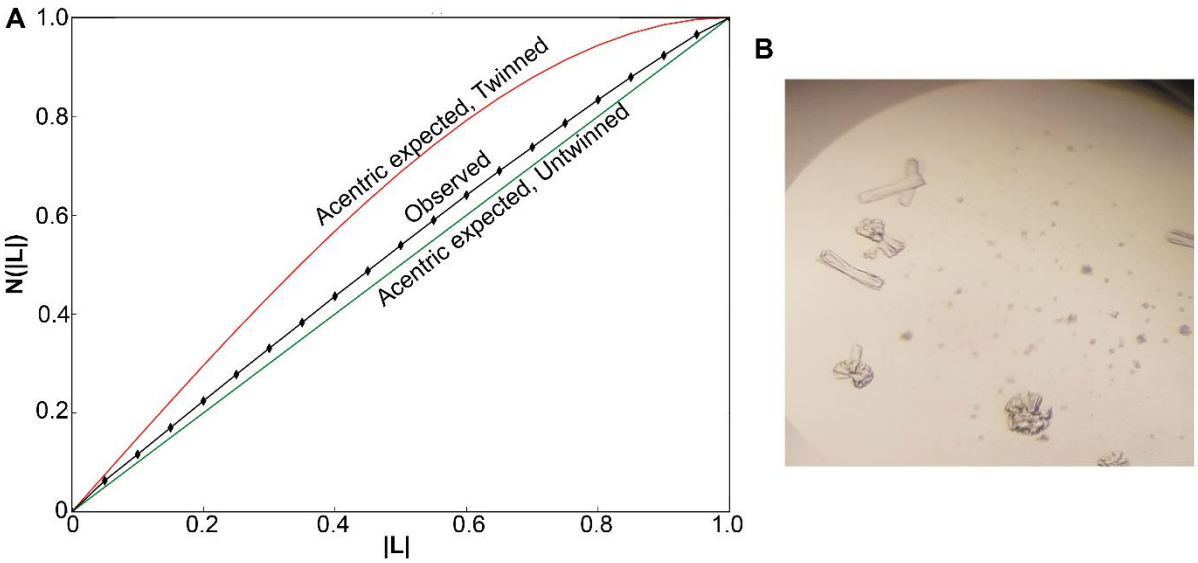



**Figure 6.9:** *L*-test for the detection of twinning in ligand free *C*Lso DHDPS. **A** is $|L|$ plotted against $N(|L|)$. $L$ is given by $(I_1-I_2)/(I_1+I_2)$, where $I_1$ and $I_2$ are the intensities of unrelated reflections and $N(|L|)$ is the cumulative probability distribution of $L$. The *L*-test suggests the data is not twinned (0.452). **B** shows *C*Lso DHDPS ligand free crystals.

XTRAIGE revealed that the likely point group of the data is P321 with the possible space groups being $P3_121$ or $P3_2$ 21. Currently, efforts towards solving the CLso ligand free structure in the two proposed space groups are ongoing.

6.2.3.2.2 *C*Lso DHDPS with ligands

The diffraction data for *C*Lso DHDPS soaked with lysine was processed in C2 spacegroup and solved by molecular replacement using the monomer of *A. tumefaciens* DHDPS as the search model (PDB ID: 4I7U, 60 % sequence identity) (Atkinson et al. 2012). The asymmetric unit contained six monomers in the asymmetric unit which corresponded to one tetramer and a dimer. Following numerous rounds of refinement, the $R_{factor}$ and $R_{free}$ were reduced to 18.9 and 22.0%, respectively.

The *C*Lso DHDPS structure with lysine was then used to solve the structure for bound pyruvate bound and for pyruvate with SSA. Following refinement, the $R_{factor}$ and $R_{free}$ were reduced to 19.0 and 0.25% for pyruvate bound and 19.0 and 23.0% for pyruvate and SSA. The solved

structure of *C*Lso DHDPS with lysine was then used in an attempt to resolve the ligand-free structure, however this was unsuccessful. All relevant data collection and structure refinement statistics are provided in Table 6.5.

T**able 6.5:** X-ray data collection and refinement statistics for *C*Lso DHDPS with bound ligands: pyruvate, SSA and pyruvate, and with lysine bound in the allosteric site. Statistical values for the highest resolution shells are given in parentheses. Each monomer within the asymmetric unit has a molecular weight of 33 kDa.

| **Data collection statistics** | **DHDPS + Pyruvate** | **DHDPS + Pyruvate and SSA** | **DHDPS + Lysine** |
|---|---|---|---|
| wavelength (Å) | 0.95369 | 0.95369 | 0.95369 |
| space group | C121 | C121 | C121 |
| unit cell parameters (Å, °) | a = 101.77, b = 133.58, c = 155.98 | a = 101.11, b = 132.92, c = 154.9, | a = 101.76, b = 133.48, c = 155.78, |
| resolution range (Å) | 46.27 – 2.4 (2.49 – 2.4) | 45.81 – 2.0 (2.07 – 2.0) | 45.2 – 2.01 (2.08 – 2.01) |
| observed reflections | 317539 (30868) | 535394 (54153) | 535866 (53764) |
| unique reflections | 79992 (7961) | 135773 (13516) | 135620 (13469) |
| mean *I/σ (I)* | 7.95 (1.04) | 9.23 (0.72) | 13.45 (1.16) |
| Completeness (%) | 99.82 (99.8) | 99.93 (99.95) | 99.91 (99.53) |
| $R_{merge}$ | 0.344 (0.776) | 0.1435 (0.898) | 0.1134 (0.553) |
| $R_{meas}$ | 0.394 (0.90) | 0.166 (1.08) | 0.131 (0.638) |
| $R\text{-}_{pim}$ | 0.191 (0.448) | 0.082 (0.514) | 0.064 (0.315) |
| CC ½ | 0.491 | 0.668 | 0.815 |
| Wilson *B*factor (Å$^2$) | 36.9 | 30.1 | 32.1 |
| Molecules per asymmetric unit | 6 | 6 | 6 |
| **Structure and refinement statistics** | | | |
| $R_{factor}$ (%) | 0.193 (0.283) | 0.197 (0.359) | 0.189 (0.293) |
| $R_{free}$ (%) | 0.249 (0.343) | 0.2 (0.3573) | 0.226 (0.325) |
| Number of atoms | | | |
| non-hydrogen | 13634 | 14208 | 14136 |
| macromolecules | 13410 | 13476 | 13464 |
| solvent | 657 | 732 | 672 |
| protein | 1770 | 1770 | 1770 |
| Root-mean-square deviation (rmsd) | | | |
| bonds (Å) | 0.008 | 0.015 | 0.016 |
| angles (°) | 1.07 | 1.87 | 1.46 |
| Ramachandran favoured | 97.64 | 97.85 | 98.24 |
| Ramachandran allowed | 2.36 | 2.15 | 1.64 |
| Ramachandran outliers | - | - | 0.11 |
| Rotamer outliers | - | - | 0.07 |
| Clashscore | 5.46 | 5.35 | 3.05 |

### *6.2.3.3 The structure of CLso DHDPS*

Consistent with the solution structure presented in Chapter Four, the crystal structure of *C*Lso DHDPS is a tetramer. The tetramer can be described as a dimer of dimers, with monomers *a* and *b* forming one dimer and monomers *c* and *d* forming the second dimer (Figure 6.10A). The monomer is composed of a $(\alpha/\beta)_8$-barrel with the active site situated deep within the centre of the β-barrel of each and a C-terminal domain consisting of three α-helices (Figure 6.10B and C). The backbone of the DHDPS tetramers of *C*Lso DHDPS and *A. tumefaciens* (PDB ID: 4I7U) align with a r.m.s.d. value of 0.26 Å indicating that they are highly similar (Figure 6.10D). *E. coli* DHDPS (PDB ID: 1YXC) aligns to both *A. tumefaciens* and *C*Lso DHDPS with a r.m.s.d of 0.88 Å, this relatively poor fit to the structures is apparent from the alignment at monomers *d* and *c* (Figure 6.10 D, E and F).

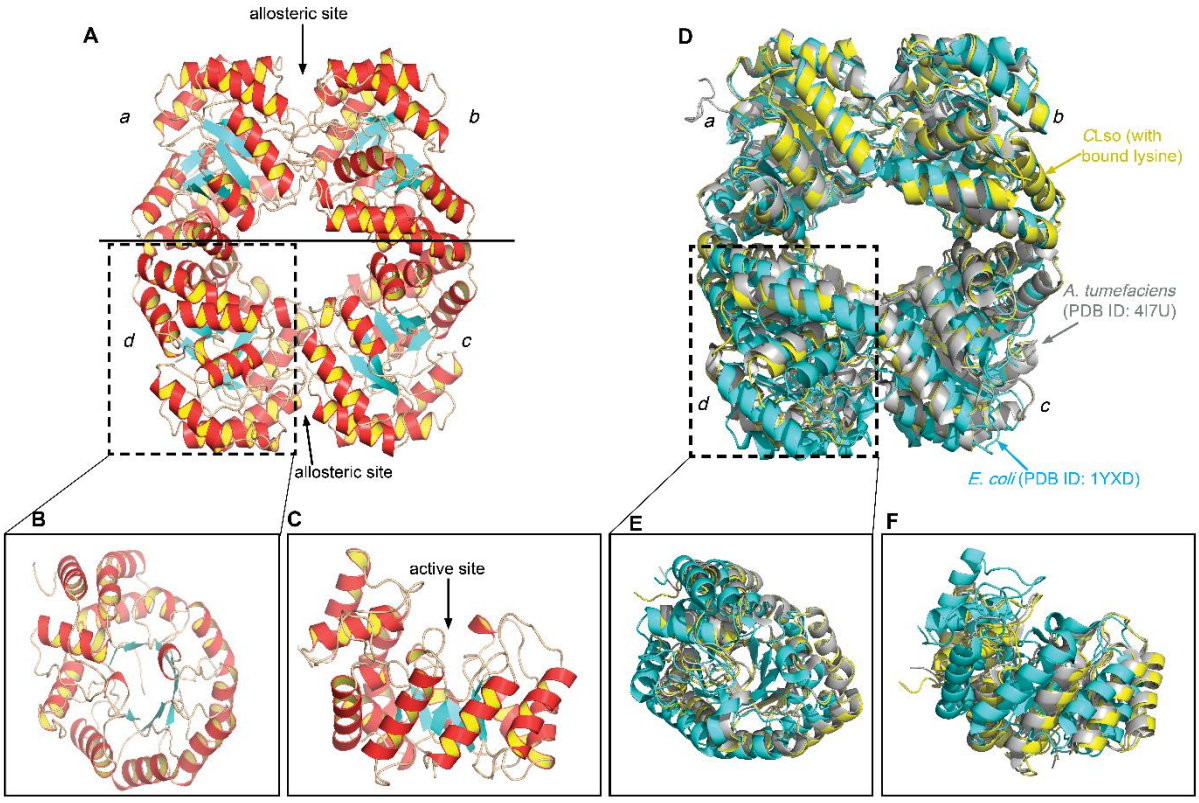



**Figure 6.10:** The structure of *C*Lso DHDPS, shown here is the structure with lysine bound in the allosteric site. **A** is the tetrameric structure of *C*Lso DHDPS with the monomers labelled *a* through *d*. The allosteric site is located between the *a/b* and *c/d* monomeric interfaces. **B** and **C** show the monomeric structure. **D** shows the overlaid structures of *C*Lso DHDPS (yellow), *A. tumefaciens* DHDPS (grey) (PDB ID: 4I7U) and *E. coli* (cyan) (PDB ID: 1YXC) with an r.m.s.d. of 0.26 and 0.88 Å respectively. **E** and **F** show the difference in alignment between for monomer *d*.

The Protein Interfaces, Surfaces and Assemblies (PISA) program available on the European Bioinformatics Institute website (Krissinel and Henrick 2007) showed that the buried surface area at the tight dimer interface of *C*Lso DHDPS is 1570 Å which is larger than that of *A. tumefaciens* (1399 Å) (Table 6.8). Furthermore, there is a more extensive dimeric interface for *C*Lso DHDPS with 22 hydrogen bonds and 5 salt bridges when compared to only 15 hydrogen bonds and 4 salt bridges in *A. tumefaciens* (Table 6.6). Comparison of the *C*Lso and *A. tumefaciens* weak dimer interface indicated that there are less contacts in *C*Lso with only 5 hydrogen bonds and 6 salt bridges compared to the 11 hydrogen bonds and 9 salt bridges in *A. tumefaciens* (Table 6.6).

**Table 6.6:** Comparison of the 'tight' dimer and 'weak' dimer interfaces of *C*Lso DHDPS and other characterised DHDPS enzymes

| Organism | Tight dimer interface (monomers *a*/*b*) | | | Weak dimer interface (monomers *b*/*c*) | | |
|---|---|---|---|---|---|---|
| | BSA[1] (Å) | Hydrogen bonds | Salt bridges | BSA (Å) | Hydrogen bonds | Salt bridges |
| *Ca*. L. solanacearum | 1570 | 22 | 5 | 943 | 5 | 6 |
| *A. tumefaciens* | 1399 | 15 | 4 | 920 | 11 | 9 |
| *E. coli* | 1289 | 9 | - | 497 | 6 | - |
| *B. anthracis* | 1392 | 10 | - | 872 | 7 | 4 |
| *T. maritima* | 1384 | 9 | - | 878 | 20 | 24 |

[1] – Buried surface area

The interface between the *a* and *b* monomers is similar to that of *A. tumefaciens*. *C*Lso DHDPS residues Tyr106 and Tyr107 are believed to be required for the interlocking of the monomers (Figure 6.11B and C). Interestingly, the *b/c* interface which connects the two dimers by residues Leu167, Thr168, Gln196 and Leu197 in *E. coli* contrast with those in *C*Lso and *A. tumefaciens*. The residues at this interface include Leu168, Glu169, Val197 and His198 for *C*Lso and Leu168, Asp69, Ala197 and His198 in *A. tumefaciens* (Figure 6.11D and E).

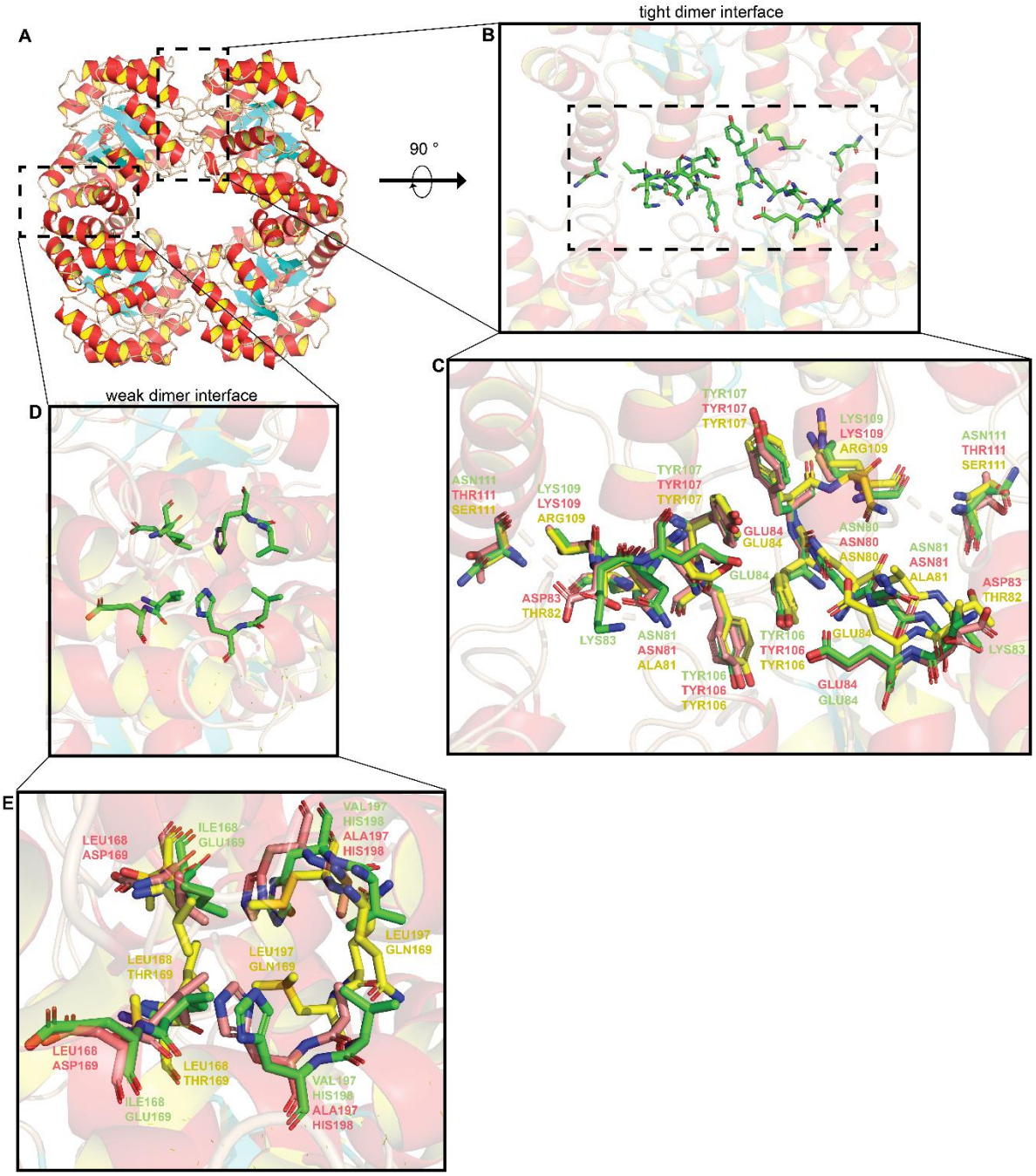



**Figure 6.11:** Key residues at the *C*Lso DHDPS interfaces. **A** is the *C*Lso DHDPS tetramer. **B** and **C** are enlargements of the tight dimer showing the key residues at the dimeric interface between monomers *a* and *b*. **D** and **E** are enlargements of the weak dimer interface between monomers *a* and *d* showing the key residues of Ile168, Glu169, Val197 and His198 at the interface. In **C** and **E** the *C*Lso interface residues (green) are overlaid with the *A. tumefaciens* residues at the interface (pink) (PDB ID: 4I7U), and with *E. coli* (yellow) (PDB ID: 1YXC).

Interestingly, the weak dimer interface has been proposed to be a target for DHDPS inhibitor design (Mitsakos et al. 2008). Mutation studies on *E. coli* and *B. anthracis* DHDPS shows that tetramerisation is essential for enzyme activity (Griffin et al. 2008; Voss et al. 2010). Further, the weak dimer interface is a region of low conservation among bacterial species. Here, the *C*Lso residues comprising the weak dimeric interface similar to *A. tumefaciens* but contrast to those in *E. coli*. The variation of residues at this weak dimer interface has allowed for the development of several compounds that display differential inhibition of DHDPS enzymes from various bacterial species (Mitsakos et al. 2008).

*6.2.3.4 The CLso DHDPS active site maintains conserved residues for catalytic activity*

The kinetic analysis demonstrated *C*Lso DHDPS binds its substrates via a ternary complex mechanism in contrast to the known mechanism for *E. coli* and *A. tumefaciens* DHDPS (ping-pong). As *C*Lso DHDPS had a much higher affinity for its substrates, it was hypothesised that, structural rearrangements in the binding site of the enzyme could account for this enhanced affinity.

Overlaying the *C*Lso active site with *A. tumefaciens* and *E. coli* DHDPS showed very little variation between the structures (Figure 6.12) and demonstrated that the catalytic triad required for activity (Thr44, Tyr107 and Tyr133) was conserved in *C*Lso. A multiple sequence alignment also demonstrates that the residues for catalytic activity (Thr44, Tyr107, Tyr133, Arg138, Gly187) were conserved between *C*Lso DHDPS and other bacterial homologues (Figure 6.13).

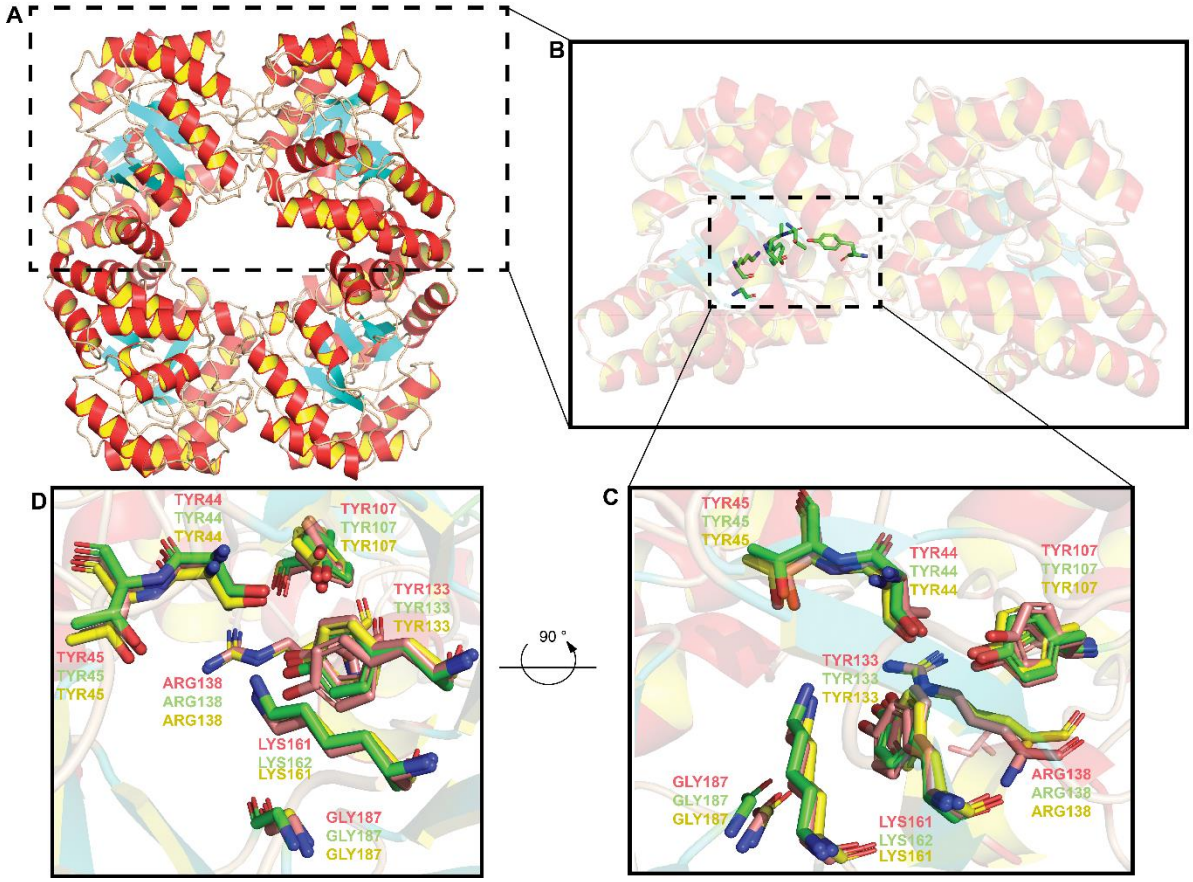



**Figure 6.12:** Key residues at the active site of *C*lso DHDPS. **A** is the *C*Lso DHDPS tetramer. **B** is an enlargement of the dimer showing the key residues of Thr44, Tyr107, Tyr133 and Lys162. **C** and **D** shows the *C*Lso interface residues (green) overlaid with the *A. tumefaciens* interface residues (pink) (PDB ID: 4I7U) and with *E. coli* (yellow) (PDB ID: 1YXC).


**Figure 6.13:** Multiple sequence alignment of *C*Lso DHDPS (KJZ81861.1) with *Thermotoga maritima* (3PB2), *B. anthracis* (KLV19457.1), *E. coli* (BAA16355.1) and *A. tumefaciens* (KJX89051.1). The green highlight indicates active site residues, while the yellow indicates allosteric site residues. Residues involved in the tight dimer interface are shown in blue and those at the weak dimer interface in red.

Similarly, there was little change in conformation of the residues with both pyruvate (Figure 6.14), and pyruvate and SSA bound (Figure 6.14) in the active site of *C*Lso DHDPS, when compared to *A. tumefaciens* DHDPS for bound pyruvate, and to *E. coli* DHDPS with bound pyruvate and SSA (PDB ID: 4EOU) (*A. tumefaciens* DHDPS with bound pyruvate and SSA not available in the PDB). Polder maps (Figure 6.14D and H) were generated for both bound pyruvate and pyruvate and SSA bound to Lys162, to ensure the substrates were modelled correctly. With no obvious structural changes in the residues required for catalytic activity, it is still unclear as to how *C*Lso DHDPS asserts such a high affinity for its substrates. Despite the *C*Lso active site having structural similarities to *A. tumefaciens* DHDPS, there is a marked contrast in substrate affinity not only to *A. tumefaciens,* but also characterised DHDPS

enzymes from *B. subtilis* (Yamakura et a. 1974), *M. tuberculosis* (Kefala et al. 2008) and *T. maritima* (Pearce et al. 2008) among others. This presents the opportunity to develop species specific inhibition of the enzyme by screening for a potential analogue of (*S*)-ASA or pyruvate.

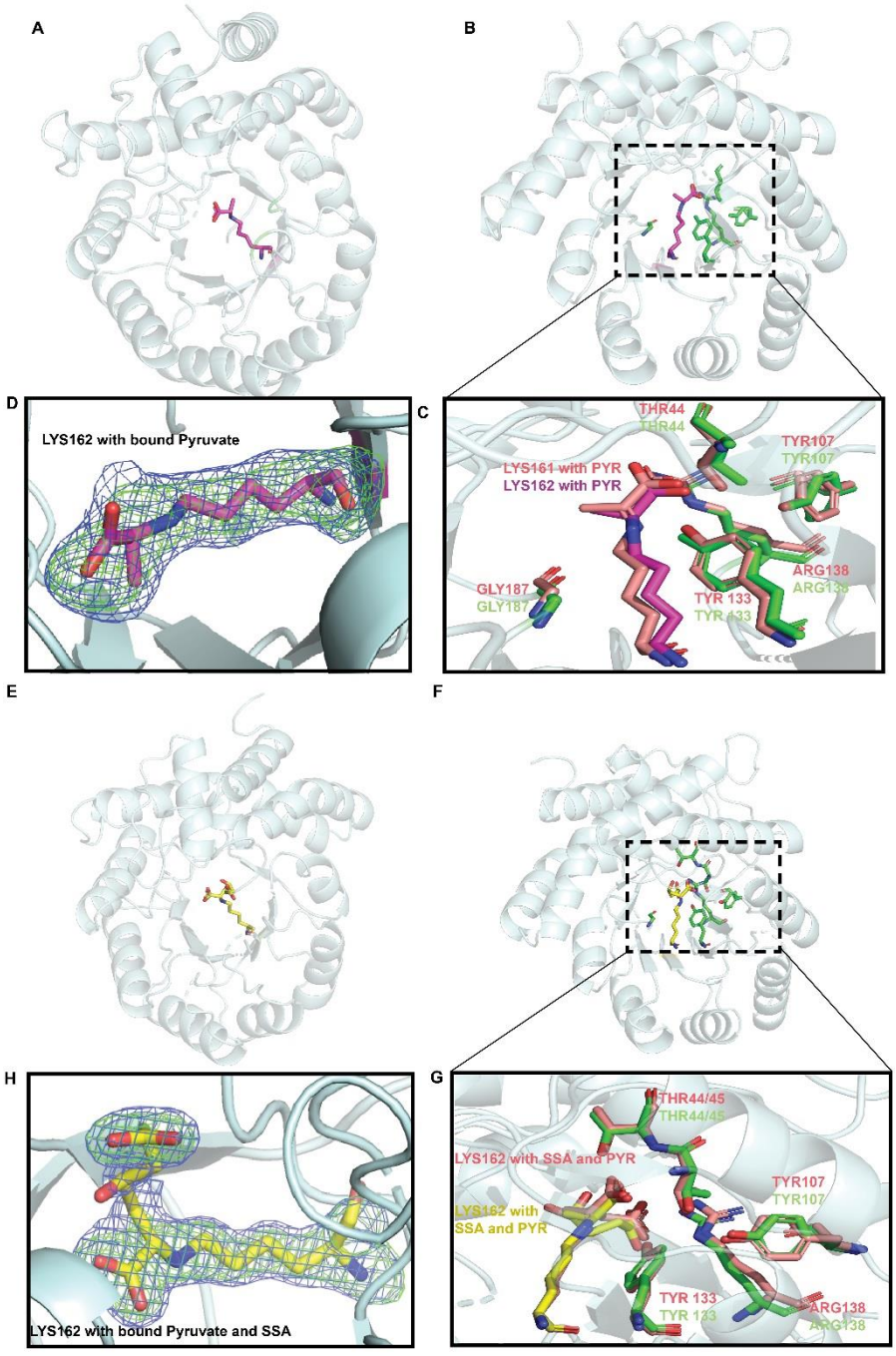



**Figure 6.14:** *C*Lso DHDPS with pyruvate, and with pyruvate and SSA bound in the active site. **A** and **E** are the *C*Lso monomer showing pyruvate and pyruvate and SSA bound to Lys162 respectively. **B** and **F** show the residues required for catalytic activity with respect to pyruvate (B) and SSA and pyruvate (F). **C** and **G** are enlargements of the active site overlaid with *A. tumefaciens* and *E. coli* respectively. **D** and **H** are polder maps showing density for pyruvate, and SSA and pyruvate in the *C*Lso DHDPS active site.

*6.2.3.5 The allosteric site*

This Chapter demonstrated that lysine binding in the allosteric site displays uncompetitive inhibition with respect to (*S*)-ASA and mixed inhibition with pyruvate, it was of interest to determine whether *C*Lso DHDPS displays a change in binding or residues involved in binding of lysine in the allosteric site. On refinement of the structure, electron density was observed between the *a* and *b* interface of the dimer. Modelling of the density supported the occupancy of two lysine molecules bound within the allosteric site of the interface between both *ab* and *cd* dimeric interfaces (Figure 6.15). On final refinement, a polder map was generated, which further supported the presence and orientation of the lysine molecules (Figure 6.15C).

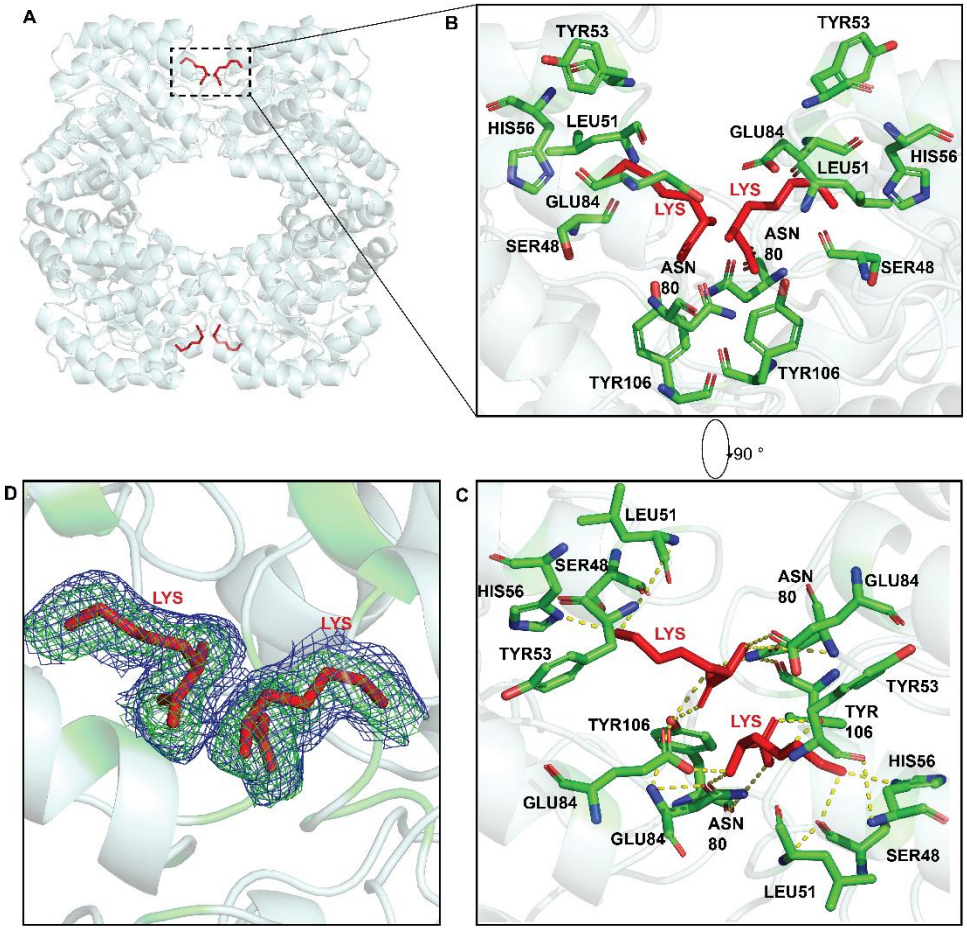



**Figure 6.15:** The allosteric site of *C*Lso DHDPS with bound lysine. **A is** the *C*Lso tetramer showing lysine (red) bound in the allosteric site between monomers *a/b* and *c/d*. **B** and **C** are enlargements of the allosteric site Polar contacts are represented by a yellow dash showing residues Leu51, Ser48, Tyr106, Asn80, Tyr 53 and His 56 within 4Å that are required for lysine binding and **D** is a polder map showing the electron density for lysine in the structure.

*C*Lso DHDPS retains these key residues for binding lysine (Figure 6.13), furthermore, there was little change in the position of the residues compared to ligand free *A. tumefaciens* (Figure 6.16). Upon lysine binding, there is a shift of Glu84 towards the lysine residue which enables a hydrogen bond to form and anchor lysine in the allosteric site (Figure 6.16B and C), this movement towards the α-amino group of lysine also occurs for *A. tumefaciens* DHDPS.

Small angle X-ray scattering was used to verify whether binding of lysine at the allosteric site induces a conformational change in *C*Lso DHDPS (Figure 6.16D). Data quality was assessed by Guinier analysis of the SAXS scattering curves at low angles (Figure 6.16D, inset) and shows that no sample aggregation or inter-particle interference is present in the data. The scattering data is presented as an intensity plot (Figure 6.16D). The experimental scatter for ligand free *C*Lso DHDPS fit the scatter for *C*Lso DHDPS with lysine with a $\chi^2$ of 0.42, indicating that there was little change in *C*Lso DHDPS structure in solution when lysine was bound in the allosteric site. However, the *P(r)* plot for the ligand free and with lysine bound *C*Lso DHDPS showed variation in the profile of the curve, this could be the result of the movement of Glu84 upon lysine binding (Figure 6.16E).

As described, the $K_i$ for *C*Lso DHDPS is 0.32 and 0.38 mM with (*S*)-ASA and pyruvate respectively, this is comparable with the $K_i$ of *E. coli* of 0.22 and 0.33 mM for pyruvate and (*S*)-ASA respectively (Dobson et al. 2004). In comparison, Atkinson et al. (2013) determined a $K_i$ of lysine of 0.063 and 0.049 mM with respect to pyruvate and (*S*)-ASA respectively for *Vitis vinefera*. The $K_i$ for lysine inhibition in plant DHDPS is much tighter than observed in bacterial DHDPS (Yugari and Gilvarg 1965; Kumpaisal et al. 1987; Burgess et al. 2008; Soares da Costa et al. 2010). Thus, an antibacterial approach targeting the allosteric site of *C*Lso DHDPS would not prove effective because of the high binding affinity for lysine in plants. Inhibitor design would subsequently have a consequence on the host plant. Thus, a future approach for *C*Lso antibacterials could be targeted to the active site rather than the allosteric site of *C*Lso DHDPS.

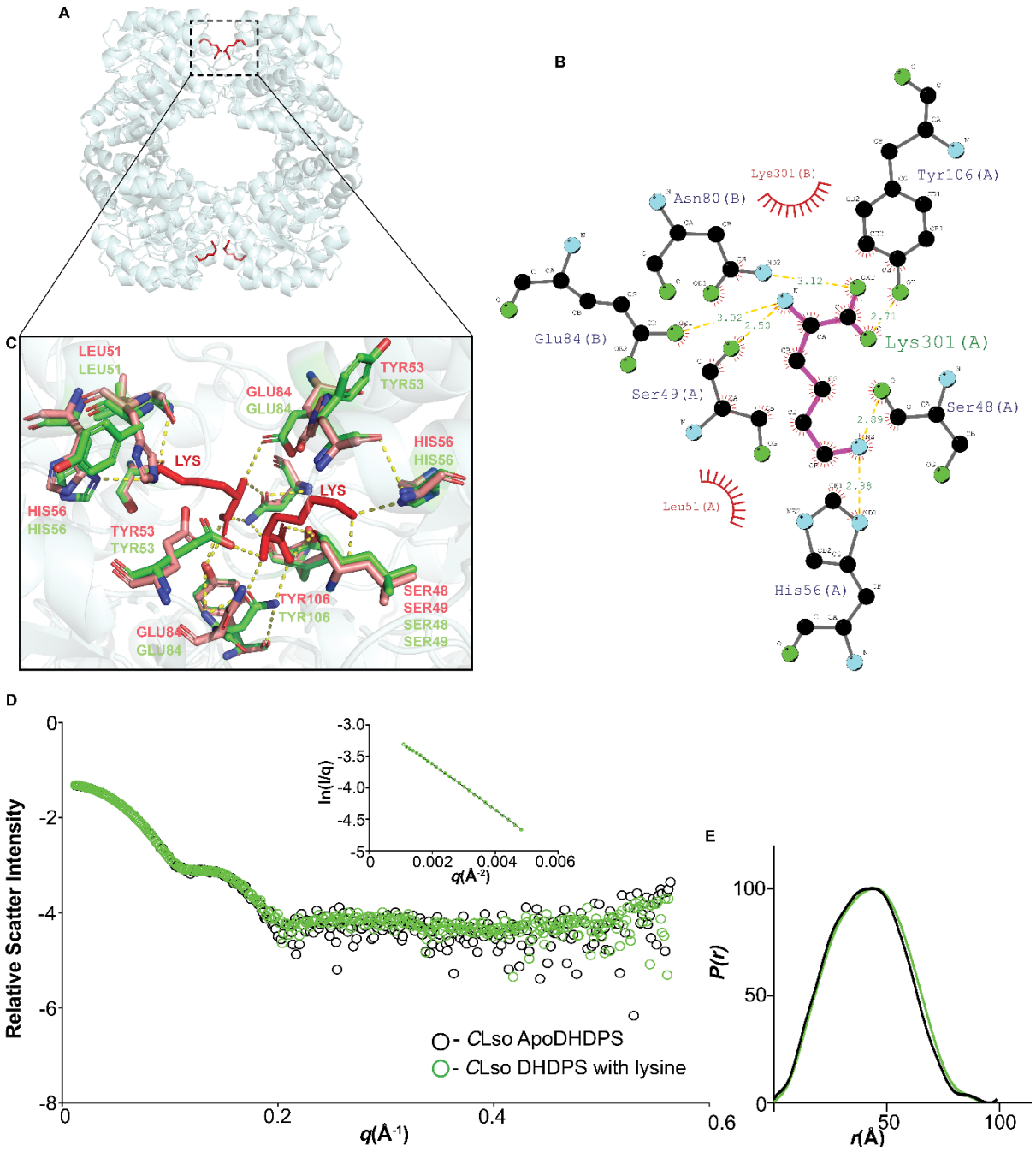



**Figure 6.16**: Analysis of changes in *C*Lso DHDPS conformation with lysine bound in the allosteric site. **A** is the *C*Lso DHDPS tetramer with **C** showing an enlargement of the allosteric site overlaid with the *C*Lso DHDPS structure with bound pyruvate to show the change in orientation of Tyr53 towards the allosteric. **B** LigPlot showing the residues required for binding of lysine. **D** is the SAXS scatter for *C*Lso DHDPS with lysine and the ligand free enzyme in solution, inset is the Guinier plot whilst **E** is the *P(r)* plot.

## 6.3 Summary and future work

Chapter Two determined that *C*Lso *dapA*, which encodes DHDPS, was overexpressed in infected psyllids compared to *in planta*. The characterisation of a functional DHDPS from *C*Lso further supports the gene expression findings in Chapter Two and has raised the question as to whether *B. cockerelli* requires lysine from intracellular organisms for survival.

This Chapter reports a detailed investigation into the structure and function of *C*Lso DHDPS, the structure for the enzyme was solved and characterised. Kinetic analysis of the enzyme reports kinetic constants that suggest a high binding affinity for substrates. It is proposed that *C*Lso DHDPS has evolved a high binding affinity for its substrates in response to nutrient limitation, which could occur in both the phloem and within the psyllid. As a result, this work provides insight into the endosymbiotic relationship between the psyllid and *C*Lso with respect to lysine biosynthesis. The active site of *C*Lso DHDPS was determined to be a potential target for antibacterial design with substrate analogues as *C*Lso DHDPS utilises a ternary-complex mechanism for substrate binding, compared to the ping-pong mechanism observed for DHDPS from other bacterial homologues.

Interestingly, lysine is not a partial inhibitor of *C*Lso DHDPS. *C*Lso demonstrates an unusual mode of lysine inhibition in that the enzyme shows mixed inhibition with respect to pyruvate and uncompetitive inhibition with respect to (*S*)-ASA, compared to the partial non-competitive model for *E. coli* DHDPS (Dobson et al. 2004). This indicates that, with respect to pyruvate, lysine binds to free DHDPS and the DHDPS-substrate complex, thus, in order to further understand the mode of lysine inhibition in *C*Lso DHDPS with respect to its substrates, future work will aim to solve the structure in complex with lysine, pyruvate and SSA.

Although there were no significant structural changes between the reduced genome *C*Lso DHDPS and *A. tumefaciens*, the change in kinetic constants highlights the necessity of structural and functional characterisation of these proteins.

## 6.4 References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66(2):213-21.

Akaike, H. 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60(2):255-65.

Ammar, E.D., Shatters, Jr. R.G., and Hall, D.G. 2011. Localization of *Candidatus* Liberibacter asiaticus, associated with citrus huanglongbing disease, in its psyllid vector using fluorescence *in situ* hybridization. Journal of Phytopathology 159(11-12):726-34.

Anderson, R.M., and May, R.M. 1982. Co-evolution of hosts and parasites. Parasitology 85(2):411-26.

Andersson, S.G., and Kurland, C.G. 1998. Reductive evolution of resident genomes. Trends in Microbiology 6(7):263-8.

Atkinson, S.C., Dogovski, C., Downton, M.T., Pearce, F.G., Reboul, C.F., Buckle, A.M., Gerrard, J.A., Dobson, R.C., Wagner, J., and Perugini, M.A. 2012. Crystal, solution and *in silico* structural studies of dihydrodipicolinate synthase from the common grapevine. PLoS One 7(6):e38318.

Atkinson, S.C., Hor, L., Dogovski, C., Dobson, R.C., and Perugini, M.A. 2014. Identification of the bona fide DHDPS from a common plant pathogen. Proteins: Structure, Function, and Bioinformatics 82(9):1869-83.

Blickling, S., Beisel, H.G., Bozic, D., Knäblein, J., Laber, B., and Huber, R. 1997. Structure of dihydrodipicolinate synthase of *Nicotiana sylvestris* reveals novel quaternary structure. Journal of Molecular Biology 274(4):608-21.

Bowden, A.C.1999. Metabolic control analysis in biotechnology and medicine. Nature Biotechnology 17(7):641.

Cicero, J.M., Fisher, T.W., Qureshi, J.A., Stansly, P.A., and Brown, J.K. 2016. Colonization and intrusive invasion of potato psyllid by '*Candidatu*s Liberibacter solanacearum'. Phytopathology 107(1):36-49.

Coulter, C.V., Gerrard, J.A., Kraunsoe, J.A., and Pratt, A.J. 1999. *Escherichia coli* dihydrodipicolinate synthase and dihydrodipicolinate reductase: kinetic and inhibition studies of two putative herbicide targets. Pesticide Science 55(9):887-95.

Cox, R.J. 1996. The DAP pathway to lysine as a target for antimicrobial agents. Natural Product Reports 13(1):29-43.

Devenish, S.R., Huisman, F.H., Parker, E.J., Hadfield, A.T., and Gerrard, J.A. 2009. Cloning and characterisation of dihydrodipicolinate synthase from the pathogen *Neisseria meningitidis*. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1794(8):1168-74.

Dijkhuizen, L. 1983. Physiological responses to nutrient limitation. Annual Review of Microbiology 37(1):1-23.

Dobson, R.C., Griffin, M.D., Roberts, S.J., and Gerrard, J.A. 2004. Dihydrodipicolinate synthase (DHDPS) from *Escherichia coli* displays partial mixed inhibition with respect to its first substrate, pyruvate. Biochimie 86(4-5):311-5.

Domigan, L.J., Scally, S.W., Fogg, M.J., Hutton, C.A., Perugini, M.A., Dobson, R.C., Muscroft-Taylor, A.C., Gerrard, J.A., and Devenish, S.R. 2009. Characterisation of dihydrodipicolinate synthase (DHDPS) from *Bacillus anthracis*. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1794(10):1510-6.

Girish, T.S., Sharma, E., and Gopal, B. 2008. Structural and functional characterization of *Staphylococcus aureus* dihydrodipicolinate synthase. FEBS letters 582(19):2923-30.

Girodeau, J.M., Agouridas, C., Masson, M., Pineau, R., and Le Goffic, F. 1986. The lysine pathway as a target for a new genera of synthetic antibacterial antibiotics?. Journal of Medicinal Chemistry 29(6):1023-30.

Griffin, M.D., Dobson, R.C., Pearce, F.G., Antonio, L., Whitten, A.E., Liew, C.K., Mackay, J.P., Trewhella, J., Jameson, G.B., Perugini, M.A., and Gerrard, J.A. 2008. Evolution of quaternary structure in a homotetrameric enzyme. Journal of Molecular Biology 380(4):691-703.

Harder, W., and Dijkhuizen, L. 1983. Physiological responses to nutrient limitation. Annual Review of Microbiology 37(1):1-23.

Hutton, C.A., Perugini, M.A., and Gerrard, J.A. 2007. Inhibition of lysine biosynthesis: an evolving antibiotic strategy. Molecular BioSystems 3(7):458-65.

Keeler, S.J., Maloney, C.L., Webber, P.Y., Patterson, C., Hirata, L.T., Falco, S.C., and Rice, J.A. 1997. Expression of de novo high-lysine α-helical coiled-coil proteins may significantly increase the accumulated levels of lysine in mature seeds of transgenic tobacco plants. Plant Molecular Biology 34(1):15-29.

Kefala, G., Evans, G.L., Griffin, M.D., Devenish, S.R., Pearce, F.G., Perugini, M.A., Gerrard, J.A., Weiss, M.S., and Dobson, R.C. 2008. Crystal structure and kinetic study of dihydrodipicolinate synthase from *Mycobacterium tuberculosis*. Biochemical Journal 411(2):351-60.

Kumpaisal, R., Hashimoto, T., and Yamada, Y. 1987. Purification and characterization of dihydrodipicolinate synthase from wheat suspension cultures. Plant Physiology 85(1):145-51.

Lin, H., Lou, B., Glynn, J.M., Doddapaneni, H., Civerolo, E.L., Chen, C., Duan, Y., Zhou, L., and Vahling, C.M. 2011. The complete genome sequence of '*Candidatus* Liberibacter solanacearum', the bacterium associated with potato zebra chip disease. PLOS One 6: e19135.

Mann, R.S., Ali, J.G., Hermann, S.L., Tiwari, S., Pelz-Stelinski, K.S., Alborn, H.T., and Stelinski, L.L. 2012. Induced release of a plant-defense volatile 'deceptively'attracts insect vectors to plants infected with a bacterial pathogen. PLoS Pathogens 8(3):e1002610.

Matthews, B.F., and Hughes, C.A. 1993. Nutritional improvement of the aspartate family of amino acids in edible crop plants. Amino Acids 4(1-2):21-34.

McCutcheon, J.P., and Moran, N.A. 2012. Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10: 13–26.

Miran, A., and Moran, N.A. 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microbial Ecology 44: 137-143.

Mitsakos, V., Dobson, R.C., Pearce, F.G., Devenish, S.R., Evans, G.L., Burgess, B.R., Perugini, M.A., Gerrard, J.A., and Hutton, C.A. 2008. Inhibiting dihydrodipicolinate synthase across species: towards specificity for pathogens. Bioorganic & Medicinal Chemistry Letters 18(2):842-4.

Nachappa, P., Levy, J., and Tamborindeguy, C. 2012. Transcriptome analyses of *Bactericera cockerelli* adults in response to '*Candidatus* Liberibacter solanacearum' infection. Molecular Genetics and Genomics 287(10): 803-817.

Pearce, F.G., Perugini, M.A., Mckerchar, H.J., and Gerrard, J.A. 2006. Dihydrodipicolinate synthase from *Thermotoga maritima*. Biochemical Journal 400(2):359-66.

Pelz-Stelinski, K.S., and Killiny, N. 2016. Better together: association with '*Candidatus* Liberibacter asiaticus' increases the reproductive fitness of its insect vector, *Diaphorina citri* (Hemiptera: Liviidae). Annals of the Entomological Society of America 109(3):371-6.

Pelz-Stelinski, K.S., Brlansky, R.H., Ebert, T.A., and Rogers, M.E. 2010. Transmission parameters for *Candidatus* Liberibacter asiaticus by Asian citrus psyllid (Hemiptera: Psyllidae). Journal of Economic Entomology 103(5):1531-41.

Sloan, D.B., and Moran, N.A. 2012 Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Molecular Biology and Evolution 29(12): 3781–3792.

Sloan, D.B., Nakabachi, A., Richards, S., Qu, J., Murali, S.C., Gibbs, R.A., and Moran, N.A. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Molecular Biology and Evolution 31(4):857-71.

Stahly, D.P. 1969. Dihydrodipicolinic acid synthase of *Bacillus licheniformis*. Biochimica et Biophysica Acta (BBA)-Enzymology 191(2):439-51.

Svergun, D.I., Barberato, C., and Koch, M.H. 1995. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. Journal of Applied Crystallography 28(6):768-73.

Tosaka, O., Takinami, K., and Hirose, Y. 1978. L-Lysine Production by S-(2-Aminoethyl) L-Cysteine and α-Amino-β-hydroxyvarelic Acid Resistant Mutants of *Brevibacterium lactofermentum*. Agricultural and Biological Chemistry 42(4):745-52.

Veldkamp, H., and Jannasch, H.W. 1972. Mixed culture studies with the chemostat. Journal of Applied Chemistry and Biotechnology 22(1):105-23.

Viola, R.E. 2001. The central enzymes of the aspartate family of amino acid biosynthesis. Accounts of Chemical Research 34(5):339-49.

Wernegreen, J.J. 2002. Genome evolution in bacterial ensosymbionts of insects. Nature Review Genetics 3:850-861.

Workneh, F., Paetzold, L., Silva, A., Johnson, C., Rashed, A., Badillo-Vargas, I., Gudmestad, N.C., and Rush, C.M. 2016. Assessments of Temporal Variations in Haplotypes of '*Candidatus* Liberibacter solanacearum' and Its Vector, the Potato Psyllid, in Potato Fields and Native Vegetation. Environmental Entomology 47(5):1184-93.

Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., and Moran, N.A. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biology 4(6):e188.

Yamakura, F., Ikeda, Y., Kimura, K., Sasakawa, T. 1974. Partial purification and some Properties of pyruvate-aspartic semialdehyde condensing enzyme from sporulating *Bacillus subtilis*. The Journal of Biochemistry 76(3):611-21.

Yang, X.B., Malik, N.S., Perez, J.L., and Liu, T.X. 2011. Impact of potato psyllid (Hemiptera: Triozidae) feeding on free amino acid composition in potato. Insect Science 18(6):663-70.

Yugari, Y., and Gilvarg, C. 1962. Coordinated end-product inhibition in lysine synthesis in *Escherichia coli*. Biochimica et biophysica acta 62(3):612-4.

Yugari, Y., and Gilvarg, C. 1965. The condensation step in diaminopimelate synthesis. Journal of Biological Chemistry 240(12):4710-6.

Zabriskie, T.M., and Jackson, M.D. 2000. Lysine biosynthesis and metabolism in fungi. Natural Product Reports 17(1):85-97.

Zwart, P.H., Grosse-Kunstleve, R.W., and Adams, P.D. 2005. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. CCP4 Newsl. 43:27-35.

# Chapter Seven

## Summary and Future Perspectives

## 7.1 Overview

The economic impact of Zebra Chip disease on the potato industry in Central and North America and New Zealand is significant. Due to the damaging effects of this disease, it is necessary to gain further insight into the biology of the causal organism, *C*Lso, because an understanding of *C*Lso pathogenicity mechanisms can inform species-specific antimicrobial drug design. Efforts to understand virulence mechanisms of Liberibacter species have been hindered by the inability to culture the bacteria in artificial media. However, strategies to control this recently emerging plant pathogen require the characterisation of correctly folded and functional proteins, which has largely been unsuccessful for the isolation of *C*Lso proteins. Efforts to understand virulence mechanisms of Liberibacter species has been hindered due to the inability to culture the bacteria in artificial media.

This thesis presents an approach to the expression and by co-expressing chaperone proteins to mimic *in vivo* cellular conditions. *C*Lso has a reduced genome and its proteins are predicted to be prone to instability and aggregation. Among intracellular-dwelling bacteria, chaperone proteins are conserved and overexpressed to buffer against problems in protein folding. With the establishment of a means for soluble *C*Lso protein expression, the thesis explored three proteins: adenyltransferase/IMP cyclohydrolase, pyruvate kinase and dihydrodipicolinate synthase in order to investigate their possible role in *C*Lso survival, metabolism and pathogenicity.

### 7.1.1 A new approach to the expression and purification of *C*Lso proteins

This thesis aimed to understand whether the effects of genome reduction alter protein stability, function and structure. One consequence of microbial genome minimisation is that, as rapid sequence evolution occurs, all genes in the genome are affected, ultimately resulting in proteins that are predicted to have less stable secondary structures (Fares et al. 2002; van Ham et al. 2003). After establishment of a symbiotic relationship, degenerative rather than adaptive genome evolution occurs (van Ham et al. 2003). The results of degenerate genome evolution include mutational bias, erosion of regulatory systems, and accumulation of

mutations affecting protein stability, continued pseudogene formation and further gene loss (van Ham et al. 2003).

Protein instability was demonstrated via the multiple expression trials of adenyltransferase/IMP cyclohydrolase with a variety of fusion tags and expression strains. Furthermore, additional *C*Lso proteins: dehydrorhamnose epimerase, dehydrorhamnose reductase, dihydrodipicolinate synthase (DHDPS), pyruvate kinase (PK) and serralysin showed protein instability by the difficulty in obtaining soluble protein with a standard *E. coli* expression system. AGGRESCAN analysis of the protein sequences indicated a higher propensity in protein aggregation propensity as genome size decreased, consistent with the theory that bacteria with reduced genomes encode unstable proteins prone to misfolding and aggregation.

This thesis is the first to successfully isolate and functionally characterise proteins from *C*Lso. Chaperone co-expression may represent a general approach for producing *C*Lso proteins in their native conformation that can then be used for further functional or structural analysis. Such studies will further our understanding of the mode of pathogenicity of the organism and assist in the design of antimicrobials against this economically significant pathogen. Future work will aim to characterise the putative pathogenicity proteins dehydrorhamnose epimerase, dehydrorhamnose reductase and serralysin that were successfully expressed but not characterised in this thesis.

### 7.1.2 The *C*Lso adenyltransferase/IMP cyclohydrolase is a monomer in solution

Adenyltransferase/IMP cyclohydrolase (ATIC) catalyses a critical step in the *de novo* purine biosynthesis pathway that is required for optimal virulence of *S. aureus*, *S. pneumoniae* (Polissi et al. 1998), *Vibrio vulnificus* (Kim et al. 2003) and *Salmonella enterica* serovar Typhimurium (McFarland and Stocker 1987). Characterisation of soluble, refolded ATIC expressed with an N-terminal $His_6$ and trigger factor fusion tag through size exclusion chromatography and analytical ultracentrifugation, highlighted the issues with present *C*Lso protein characterisation studies that use a denaturation and refolding. This Chapter showed that although soluble, $His_6$TF-*C*LsoATIC was composed of many higher oligomeric structures in solution. Chaperone co-expression of *C*Lso ATIC resulted in folded protein; however, no

activity was observed. One caveat to the work was the lack of a positive control to confidently determine that *C*Lso ATIC was not functional. Interestingly, Lin et al. (2011) found that *C*Lso lack several of the alternative routes for nucleotide synthesis found in closely-related Rhizobiaceae. This is consistent with the highly specialised lifestyle and reduced genome of Liberibacter species. Thus, I speculate that with no observable ATIC activity and reduced mechanisms for the synthesis of nucleotides, it raises the question as to whether *C*Lso encodes a functional *de novo* purine biosynthesis pathway. Future work will aim to investigate whether *C*Lso ATIC is capable of restoring purine synthesis in *E. coli* knockout strains void of *purH*. In addition, I aim to isolate the *E. coli* recombinant ATIC enzyme to serve as a positive control for future assessment of *C*Lso ATIC functionality.

### 7.1.3 *C*Lso pyruvate kinase has a structural similarities to the *E. coli* enzyme with a notable change in a fructose-1,6-bisphosphate binding residue.

A structural investigation into *C*Lso pyruvate kinase (PK) is discussed. Interestingly, in *E. coli* Lys382 is required for binding of the allosteric activator, fructose-1,6-bisphosphate whereas *C*Lso has Gln383 at the same position. Lys382 was shown to prevent transition to the R-state and affect functionality in *E. coli*. This work showed that Gln383 is situated in an optimal position to bind fructose-1,6-bisphosphate however, future work will aim to kinetically characterise *C*Lso PK with respect to each of the substrates, phosphoenolpyruvate and ADP, in the presence and absence of fructose-1,6-bisphosphate to determine whether *C*Lso PK is allosterically activated. Furthermore, investigation into *C*Lso Gln383 to the Lys to determine whether the glutamine residue effects fructose-1,6-bisphosphate binding and activity of the enzyme.

### 7.1.4 *C*Lso dihydrodipicolinate synthase binds substrates via a ternary complex mechanism and is not partially inhibited by lysine.

A detailed structural and functional investigation of *C*Lso DHDPS synthase is reported. Although there were no significant structural differences between the *Agrobacterium tumefaciens* and *C*Lso active and allosteric sites, kinetic analysis showed considerable differences. *C*Lso DHDPS binds substrates via a ternary complex mechanism, the first example of this mechanism for a DHDPS enzyme. In addition, *C*Lso demonstrates an unusual mode of

lysine inhibition in that the enzyme shows mixed inhibition with respect to pyruvate and uncompetitive inhibition with respect to (*S*)-aspartate semialdehyde ((*S*)-ASA), compared to the partial non-competitive model for *E. coli* DHDPS (Dobson et al. 2004).

The kinetic constants reported here for *C*Lso DHDPS are the first for an enzyme from the organism. The values obtained show that *C*Lso dihydrodipicolinate synthase has a higher binding affinity for the substrates pyruvate and (*S*)-ASA. I propose that this may be the result of *C*Lso adaptation to a nutrient-limited environment within the psyllid host and *in planta*. Nutrient-limited environments exert selective pressures that allow the organism to evolve mechanisms to survive within the restrictive environment (Harder and Dijkhuizen 1983). I hypothesise that the increase in binding affinity, and the increase in expression of *dapA* in psyllids, provide the organism with a mechanism to increase the rate of initial metabolism of the substrates accumulate inside the cell when the intracellular concentration is low, deriving a high-affinity enzyme system (Harder and Dijkhuizen 1983). This work provides insight into the endosymbiotic relationship between the psyllid and *C*Lso with respect to lysine biosynthesis and may indicate how proteins within the *C*Lso adapt to nutrient limitations within the psyllid and *in planta*. Future work will aim to structurally investigate changes in substrate binding in the presence of lysine by solving the structure of *C*Lso DHDPS in complex with lysine and pyruvate, and with pyruvate, (*S*)-ASA and lysine.

## 7.2 Summary

To date, *C*Lso biology and pathogenicity has been inferred based on genome annotation (Lin et al. 2011). However, this thesis demonstrates that protein function cannot be predicted based on sequence analysis as the intracellular lifestyle of *C*Lso has resulted in changes in protein function, mode of substrate binding and inhibition. Genetic drift occurs as a result of frequent bottlenecks in bacterial transmission associated with the intracellular and is believed to alter protein function. The resulting effect results in the fixation of mutations that inactivate genes that are useful, as demonstrated with the *de novo* purine synthesis protein adenyltransferase/IMP cyclohydrolase. Nutrient limitation within the psyllid host and in the phloem of plants is believed to exert a selective pressure and result in adaptation of *C*Lso proteins to these environments. As demonstrated by dihydrodipicolinate synthase,

environmental adaptation has resulted in a DHDPS enzyme uniquely different to those previously characterised. Therefore, this thesis shows the need for future characterisation *C*Lso proteins as they could demonstrate unique mechanisms for substrate binding or inhibition distinct from other bacterial homologues. Thus, changes in *C*Lso protein characteristics could provide a unique target for species specific antibacterials.

## 7.3 References

Dobson, R.C., Griffin, M.D., Roberts, S.J., and Gerrard, J.A. 2004. Dihydrodipicolinate synthase (DHDPS) from *Escherichia coli* displays partial mixed inhibition with respect to its first substrate, pyruvate. Biochimie 86(4-5):311-5.

Fares, M.A., Ruiz-González, M., Moya, A., Santiago, E.F., and Barrio, E. 2002. Endosymbiotic bacteria: *groEL* buffers against deleterious mutations. Nature 417:398.

Harder, W., and Dijkhuizen, L. 1983. Physiological responses to nutrient limitation. Annual Review of Microbiology 37(1):1-23.

Kim, J.K., Am Jang, H., Won, Y.J., Kikuchi, Y., Han, S.H., Kim, C.H., Nikoh, N., Fukatsu, T., and Lee, B.L. 2014. Purine biosynthesis-deficient *Burkholderia* mutants are incapable of symbiotic accommodation in the stinkbug. The ISME journal 8(3):552.

Lin, H., Lou, B., Glynn, J.M., Doddapaneni, H., Civerolo, E.L., Chen, C., Duan, Y., Zhou, L., and Vahling, C.M. 2011. The complete genome sequence of '*Candidatus* Liberibacter solanacearum', the bacterium associated with potato zebra chip disease. PLOS One 6: e19135.

McFarland, W.C., and Stocker, B.A. 1987. Effect of different purine auxotrophic mutations on mouse-virulence of a Vi-positive strain of *Salmonella dublin* and of two strains of *Salmonella typhimurium*. Microbial Pathogenesis 3(2):129-41.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. 1998. Large-Scale Identification of Virulence Genes from *Streptococcus pneumoniae*. Infection and Immunity 66(12):5620-9.

van Ham, R.C.H.J., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J.M., Jimenez, L., Postigo, M., Silva, F.J., Tamames, J., Viguera, E., Latorre, A., Valencia, A., Moran, F., and Moya, A. 2003 Reductive genome evolution in *Buchnera aphidicola*. Proceedings of the National Academy of Sciences of the United States of America 100: 581–586.

# Chapter Eight

## Materials and Methods

## 8.1 Experimental reagents

### 8.1.1 Chemical reagents

Unless otherwise stated, general chemicals, including antibiotics, were purchased from Sigma-Aldrich (St. Louis, MO). Chemicals used to prepare growth media, including Luria Bertani broth base, were purchased from Invitrogen (Waltham, MA), while agar was obtained from Oxoid (Hampshire, United Kingdom). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was obtained from Bioline (London, United Kingdom).

### 8.1.2 Molecular, biological and biochemical reagents

The Aurum total RNA fatty and fibrous tissue kit for RNA extraction from the tomato potato psyllid was obtained from BioRad (Hercules, CA) and the Geneaid (New Taipei City, Taiwan) total RNA mini kit was used for RNA extraction from plant samples. The iScript™ advanced cDNA synthesis kit for RT-qPCR and SsoAdvanced™ universal SYBR® Green Supermix were both purchased from BioRad. RNAlater was purchased from Thermo Fisher Scientific (Waltham, MA).

Plasmid constructs harbouring genes of interest were purchased from GenScript (Piscataway, NJ). DNA oligonucleotides, primers and any associated chemical modification were purchased from Integrated DNA Technologies (IDT) (San Jose, CA). All restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Agarose gel DNA extraction kits were supplied by Roche (Basel, Switzerland). DNA-Spin™ Plasmid DNA Purification Kits (miniprep) were obtained from iNtRon Biotechnology (Gyeonggi-do, Korea). T4 DNA Ligase and In-Fusion® HD cloning kits, including chemically competent Stellar *E. coli* cells were purchased from TaKaRa Bio (Shiga Prefecture, Japan). The pOPIN vector suite was obtained from Addgene (Watertown, MA) and Tuner™ (DE3)pLacI competent *E. coli* cells were obtained from Novagen (Birmingham, United Kingdom). The chaperone plasmid set was purchased from TaKaRa Bio. Chemically competent *E. coli* BL21(DE3) cells were obtained from New England Biolabs.

10-formyl-tetrahydrofolate and 5-formylaminoimidazole-4-carboxamide ribonucleotide were generously synthesised by Zoe Wright of the Wolan Laboratory at the Scripps Research

Institute, San Diego, CA. Bradford dye reagent was obtained from BioRad. DNA HyperLadder™ marker was supplied by Bioline. Novex® Sharp Pre-stained Protein Ladder, Novex® Tris-Glycine SDS sample buffer, Benchmark™ Protein Ladder and BOLT™ running buffer were purchased from ThermoFisher. SYBR® Safe DNA gel stain and SYPRO® Orange dye were also obtained from ThermoFisher.

### 8.1.3 General consumables

All prepacked chromatography columns were purchased in various sizes from GE Healthcare. Syringe filters (0.22- and 0.45-µm) and microcentrifuge tube spin concentrators (various molecular weight cut-offs) were purchased from Merck Millipore (Burlington, MA). Crystallisation screens, sitting-drop crystallisation plates, UV-transmissible plate seals and cryo-protection solutions were purchased from Molecular Dimensions (Suffolk, UK). Cryo loops were obtained from Hampton Research (Journey Aliso Viejo, CA).

## 8.2 General methodology

### 8.2.1 Sequencing

The sequence of cloned constructs was confirmed using Sanger sequencing with appropriate primers (Macrogen, Seoul, South Korea). Sequencing results were aligned with reference sequences using Geneious (Biomatters, Newark, NJ).

### 8.2.2 Centrifugation

Centrifugation was performed using either a Sorvall LYNX 4000 Superspeed (Thermo Fisher), a Sorvall RC 6 Plus (Thermo Fisher) or an Eppendorf 5430 R centrifuge.

### 8.2.3 pH measurement

Buffer pH was measured using an Ultrabasic Benchtop pH meter (Sartorius, Goettingen Germany) calibrated using appropriate pH standards. The pH of solutions was then adjusted using 1 M or 10 M HCl or NaOH.

### 8.2.4 Sterilisation

Media and glassware for bacterial cultures sterilised by autoclaving at 121°C for 15 minutes. Solutions were prepared with MilliQ water (MQW). Aseptic technique was used for each experiment to prevent contamination.

### 8.2.5 Antibiotics

Antibiotics were prepared as 1000× stock solutions and stored at -20°C. The working concentrations used in this thesis are listed in Table 8.1.

**Table 8.1:** Antibiotics used in this thesis.

| Antibiotic | Stock Concentration (mg/mL) | Working Concentration (µg/mL) | Solvent |
|---|---|---|---|
| Ampicillin | 100 | 100 | MQW |
| Kanamycin | 50 | 50 | MQW |
| Chloramphenicol | 34 | 34 | Ethanol |
| Tetracycline | 12.5 | 12.5 | Methanol |

## 8.3 Microbiological methods

### 8.3.1 Bacterial strains

The strains of *Escherichia coli* used for expression of each of the proteins in this research are described in Table 8.2.

**Table 8.2:** Bacterial strains used in this thesis. The strain name and associated genotype is presented.

| Protein | *E. coli* strain | Genotype |
|---|---|---|
| Adenyltransferase/IMP cyclohydrolase | Tuner(DE3) pLacI | F– *ompT hsdSB (rB– mB–) gal dcm lacY1*(DE3) pLacI (CmR) |
| | BL21(DE3) | *fhuA2 [lon] ompT gal* (λ DE3) *[dcm] ΔhsdS* |
| | BL21(DE3) pLysS | F–*, ompT, hsdSB (rB–, mB–), dcm, gal,* λ(DE3)*,* pLysS*,* CmR |
| | Origami | F– *ompT hsdSB(rB- mB-) gal dcm lacY1 ahpC (DE3) gor522:: Tn10 trxB (*KanR*,* TetR*)* |
| | C41(DE3) | F– *ompT hsdSB (rB- mB-) gal dcm* (DE3) |
| | C43(DE3) | F– *ompT hsdSB (rB- mB-) gal dcm* (DE3) |
| | C41(DE3) pLysS | F– *ompT hsdSB (rB- mB-) gal dcm* (DE3) pLysS (CmR) |
| | C43(DE3) pLysS | F– *ompT hsdSB (rB- mB-) gal dcm* (DE3) pLysS (CmR) |
| Dihydrodipicolinate synthase<br>Pyruvate kinase<br>Dehydrorhamnose epimerase<br>Dehydrorhamnose reductase<br>serralysin<br>Dehydrodipicolinate reductase | BL21(DE3) | F– *fhuA2 [lon] ompT gal* (λ DE3) *[dcm] ΔhsdS* |
| | Stellar[1] | F–*, endA1, supE44, thi-1, recA1*, *relA1*, *gyrA96, phoA, Φ80d lacZΔ M15, Δ (lacZYA-argF) U169, Δ* (mrr-*hsdRMS-mcrBC*), *ΔmcrA, λ–* |

[1]Cloning strain of competent cells used for each of the proteins

### 8.3.2 Plasmid constructs

The constructs used for expression of *C*Lso adenyltransferase/IMP cyclohydrolase are described in Table 8.3.

**Table 8.3:** Constructs used for the expression of *C*Lso adenyltransferase/IMP cyclohydrolase

| Cloning Vector | *E. coli* Strain | Application | Protein construct |
|---|---|---|---|
| pET28a | BL21(DE3)<br>BL21(DE3) pLysS<br>Origami<br>C41(DE3)<br>C43(DE3)<br>C41(DE3) pLysS<br>C43(DE3) pLysS | Protein solubilisation trials | $His_6$-ATIC |
| pET28a | BL21(DE3) | Separate ATIC domains | $His_6$-AICARFT[1] |
| pET28a | BL21(DE3) | Separate ATIC domains | $His_6$-IMPCH[2] |
| pET24a | BL21(DE3)[3] | Chaperone co-expression | ATIC-$His_6$ |
| pOPIN[4] | Tuner(DE3) pLacI | Fusion tag solubilisation trials | $His_6$-Fusion-ATIC |

[1]AICARFT, 5-amino-4-imidazolecarboxamide ribonucleotide formyl transferase domain
[2]IMPCH, inosine monophosphate cyclohydrolase domain
[3]Expressed in BL21(DE3) containing plasmids for the co-expression of chaperone proteins detailed in Table 8.4.
[4]pOPIN expression vector details are presented in Table 8.4.

#### *8.3.2.1 pOPIN fusion protein expression constructs*

Vectors from the pOPIN suite were a gift from Ray Owens (Berrow et al. 2007). *C*Lso *pykF, dapA, znMC, rmlC and rmlD* were cloned into the pOPINF vector whilst *purH* was trialled with each pOPIN construct in order to aid in protein solubility (Addgene plasmid # 26042, a gift from Ray Owens; Berrow et al. 2007) to generate expression vectors with a general structure of N-$His_6$-HRV3C protease cleavage site-protein of interest expression of *C*Lso adenyltransferase/IMP cyclohydrolase are listed in Table 8.4. The pOPIN vectors were transformed into Tuner(DE3) pLacI competent cells.

**Table 8.4:** pOPIN constructs used for the expression of *C*Lso adenyltransferase/IMP cyclohydrolase expressed in Tuner(DE3) pLacI.

| pOPIN Vector | Addgene ID | Product | Molecular Weight (kDa) | Fusion Tag |
|---|---|---|---|---|
| F | 26042 | $His_6$-3C[1]-ATIC | 61 | N-$His_6$ |
| S3C | 41115 | $His_6$-SUMO-3C-ATIC | 72 | N-$His_6$ SUMO |
| M | 26044 | $His_6$-MBP-3C-ATIC | 73 | N-$His_6$ maltose binding protein |
| TRX | 41116 | $His_6$-TRX-3C-ATIC | 75 | N-$His_6$ thioredoxin |
| J | 26045 | $His_6$-GST-3C-ATIC | 87 | N-$His_6$ GST |
| TF | 41114 | $His_6$-TF-3C-ATIC | 87 | N-$His_6$ trigger factor |
| HALO7 | 41117 | $His_6$-HALO7-3C-ATIC | 95 | N-$His_6$ HALO |
| MSYB | 41118 | $His_6$-MSYB-3C-ATIC | 102 | N-$His_6$ hyper acidic protein fusion |
| NUSA | 41119 | $His_6$-NUSA-3C-ATIC | 109 | N-$His_6$ N-utilization substance |
| GFP | 53541 | $His_6$-NGFP-3C-ATIC | 116 | N-$His_6$ green fluorescent protein |

[1]3C cleavage site for the human rhinovirus 3C protease (HRV 3C Protease) that recognizes the cleavage site of Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro

### *8.3.2.2 Constructs for chaperone expression*

Plasmids used for chaperone co-expression are detailed in Table 8.5. *C*Lso *pykF, dapA, znMC, rmlC, rmlD* and *purH* were codon-optimised for expression in *E. coli* then synthesised and cloned into the pET24a expression plasmid by GenScript. The pET24a plasmids containing the *C*Lso gene of interest were transformed into *E. coli* BL21(DE3) strains that harboured plasmids encoding chaperones for co-expression (Takara Bio Inc, Shiga, Japan) (Table 8.5).

**Table 8.5:** Plasmids used for co-expression of chaperones with *C*Lso proteins

| Plasmid | Chaperone | Expression Host | Chaperone molecular weight (kDa) | Promoter | Inducer |
|---|---|---|---|---|---|
| pGRO7 | GroES<br>GroEL | *E.coli* BL21(DE3) | 10<br>60 | *araB* | L-arabinose |
| pKJE7 | dnaK<br>dnaJ<br>grpE | *E.coli* BL21(DE3) | 70<br>40<br>22 | *araB* | L-arabinose |
| pTf16 | Trigger factor | *E.coli* BL21(DE3) | 56 | *araB* | L-arabinose |

### 8.3.3 Media

*8.3.3.1 Luria Bertani medium*

Luria Bertani (LB) broth (LB-Lennox: 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) was purchased as a ready-to-use powder from ThermoFisher. For 1 L of MilliQ water, 20 g of LB base was added. The resulting medium was sterilised by autoclaving and stored for later use at room temperature.

*8.3.3.2 LB agar*

LB agar was prepared using MilliQ water. For every 1 L of MilliQ water, 20 g of LB base and 12 g of agar were added and the medium sterilised via autoclaving. The molten media was supplemented with the required antibiotic and mixed by swirling. The medium was then poured into sterile petri dishes in a laminar flow hood. Plates were left to set under sterile conditions before being sealed with parafilm and stored at 4°C for later use.

### 8.3.4 Preparation of competent cells

Chemically competent *E. coli* cells were initially obtained from commercial sources. Cloning and expression strains used routinely were prepared in-house using a calcium chloride-based method (Sambrook and Russel 2006). Commercial stocks were streaked onto LB agar plates

without antibiotic and incubated overnight at 37°C. Following overnight incubation, a single colony was inoculated into 10 mL of sterile LB medium and incubated overnight at 37°C, without antibiotic. For pLysS and pLacI strains, a single colony was inoculated into 10 mL of sterile LB medium and incubated overnight at 37°C with chloramphenicol. From this overnight culture, 1 mL of starter culture was added to 100 mL of LB and left to grow at 37°C until an $OD_{600}$ of 0.35–0.4. Cells were immediately put on ice and cooled for 30 min. Once cool, cells were centrifuged for 5 min at 13,000 × g. The supernatant was decanted and resuspended in 10 mL of sterile, ice cold 0.1 M $CaCl_2$ before a further 30 min incubation on ice. Cells were then centrifuged for 5 min at 13,000 × g. The supernatant was discarded, and cells were resuspended in 5 mL of sterile, ice cold 0.1 M $CaCl_2$ with 15% (v/v) glycerol. Chemically competent cells were aliquoted into sterile 1.7-mL micro-centrifuge tubes under aseptic conditions, flash-frozen and stored at -80°C until further use.

*8.3.4.1 Transformation of chemically competent bacterial cells*

Competent cells were thawed on ice prior to transformation. Approximately 100 ng of DNA was added to an aliquot of competent cells and incubated on ice for 20 min, followed by heat shock at 42°C for 45 s. Cells were returned to ice for 5 min to cool. Subsequently, 400 µL of SOC media was added and cells were incubated at 37°C with agitation for 1 h. Aliquots were plated onto LB agar plates with appropriate antibiotic selection and incubated at 37°C overnight.

**8.3.5 Glycerol stocks**

A single transformed colony was selected to inoculate a 10-mL LB starter culture, supplemented with appropriate antibiotics then incubated at 37°C overnight. An aliquot was added to a fresh 10-mL LB culture and left to grow at 37°C, with shaking at 220 rpm, until an $OD_{600}$ of 0.4–0.5 was reached. An aliquot of culture was then added to a sterile 1.7-mL micro-centrifuge tube and glycerol was added to final concentration of 15% (v/v). Tubes were subsequently flash-frozen in liquid nitrogen and stored at -80°C until further use.

## 8.4 Molecular biology and biochemistry

### 8.4.1 DNA preparation

Double-stranded linear DNA fragments encoding *dapA, purH, rmlC, rmlD, znMC* and *pykF* were synthesised as gBLOCKS (IDT, San Jose, CA) for parallel cloning into the pOPIN vectors. Double stranded DNA (dsDNA) oligonucleotides were stored at -20 °C until use.

### 8.4.2 Plasmid preparation

Plasmid DNA was purified using a DNA-Spin™ Plasmid DNA Purification Kit (iNtRon Biotechnology) according the manufacturer's instructions. Briefly, plasmids were transformed into chemically competent *E. coli* cells (Stellar strain) by heat shock. A single transformed colony was selected to inoculate a 10-mL LB starter culture and was incubated at 37°C overnight. Following incubation, the cells were centrifuged, and plasmid DNA was purified according to the manufacturer's instructions. The DNA sequence of plasmids was verified by Sanger sequencing.

### 8.4.3 InFusion cloning

InFusion cloning is a ligation-independent cloning method and was performed as per the manufacturer's instructions (TaKaRa Bio USA). This process was used for the cloning of synthesised gBLOCKS encoding *dapA, purH, rmlC, rmlD, znMC* and *pykF* into the pOPIN plasmid. The process is explained below.

#### *8.4.3.1 Restriction enzyme digestion*

pOPIN vectors were digested with the *Kpn*I and *Hind*III restriction enzymes (New England BioLabs) to linearise the plasmid. The components of the restriction double digest are shown in Table 8.6, where reactions were incubated for 3 h at 37 °C.

**Table 8.6:** Reaction parameters for a restriction double digest.

| Reagent | Amount |
|---|---|
| Plasmid DNA | Approximately 2000 ng DNA |
| *Kpn*I | 1 µL |
| *Hind*III | 1 µL |
| 10× CutSmart® buffer | 5 µL |
| Nuclease-free MilliQ water | x µL |
| *Total volume* | 50 µL |

*8.4.3.2 DNA gel extraction*

Following restriction enzyme digest the reaction mixture was separated by agarose gel electrophoresis. DNA bands with a consistent size of the pOPIN vector band of interest were excised from the gel using a razor blade and extracted from the gel using an Agarose Gel DNA Extraction kit, according to the manufacturer's instructions (Roche). Synthesised gBLOCKs were then InFusion cloned into the empty pOPIN plasmids.

*8.4.3.3 Polymerase Chain Reaction*

Polymerase chain reaction (PCR) was performed using CloneAmp HiFi PCR Premix, according to manufacturer's instructions (TaKaRa Bio USA). Briefly, the PCR reaction contained 1 µL of CloneAmp HiFi PCR Premix, 1 µL of the synthesised gBLOCK gene fragment, 1 µL of digested pOPIN plasmid and 2 µL of nuclease-free MilliQ water (Thermo Fisher). The PCR amplification was performed as per conditions in Table 8.7 using a Verti 96-well thermal cycler (Applied Biosystems). Plasmids were purified using a NucleoSpin Plasmid Miniprep kit (Macherey-Nagel, Düren, Germany) and DNA sequence was verified by Sanger sequencing (Macrogen, Seoul, Korea).

**Table 8.7:** PCR amplification parameters

| Temperature (°C) | Time (s) | Cycle |
|---|---|---|
| 98 | 30 | 1 |
| 98 | 10 | |
| 50–65 °C [1] | 15 | 30 |
| 72 | 30–60 per kb | |
| 72 | 120 | 1 |
| 4 | ∞ | 1 |

[1]Annealing temperature is primer dependent

### 8.4.4 Determination of protein and DNA concentrations

#### *8.4.4.1 Bradford protein assay*

Protein concentration was measured using the Bradford assay (Bradford 1976). To prepare a 1× dye solution, one part of concentrated Bradford dye reagent was diluted in four parts MilliQ water and left to equilibrate at room temperature. A calibration curve was obtained by serially diluting bovine serum from 0.5 mg/mL to 0.016 mg/mL in the same buffer that the protein sample was in. Protein samples (10 µL) were mixed with 200 µL of 1× Bradford dye reagent and added to a 96-well microplate in triplicate. A blank was prepared using 10 µL of MilliQ water instead of protein. Samples were incubated for 5 min at room temperature prior to measurement of the absorbance at 595 nm using a SpectraMax M5 microplate spectrophotometer (Molecular Devices, Sunnyvale, CA). Unknown protein samples were diluted in accordance with the midpoint of the calibration curve, where the concentration was extrapolated from the absorbance at 595 nm.

#### *8.4.4.2 Quantitation of proteins using ultraviolet spectroscopy*

The concentration of protein was also determined by ultraviolet (UV) spectrometry. UV absorbance at a wavelength of 280 nm was measured using a NanoDrop™ 1000 spectrometer (Thermo Fisher). The extinction coefficient for each protein was predicted from the sequence using ExPASy ProtParam (https://web.expasy.org/protparam/) based on the content of the aromatic residues, tryptophan and tyrosine.

*8.4.4.3 Quantitation of DNA using ultra-violet spectroscopy*

The concentration of DNA was determined using the Nucleic Acid module of a NanoDrop™ 1000 spectrometer. UV absorbance at a wavelength of 260 nm.

## 8.5 Electrophoresis

### 8.5.1 Electrophoresis buffers and solutions

Buffer solutions used for either agarose gel or sodium dodecyl sulphate polyacrylamide gel electrophoresis are listed in Table 8.8.

**Table 8.8:** Buffers and solutions for gel electrophoresis

| Solution | Composition |
|---|---|
| 1× Tris-acetate-EDTA (TAE) buffer | 40 mM Tris-HCl (pH 8.0), 0.11 % (v/v) glacial acetic acid, and 1 mM EDTA |
| 6× DNA loading buffer | 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, and 30% (v/v) glycerol |
| 20× Bolt™ MES SDS running buffer | 50 mM 2-(N-morpholino)ethanesulfonic acid (MES), 50 mM Tris-HCl (pH 7.3), 0.1% SDS, and 1 mM EDTA |
| 1× Tris-Glycine SDS Running Buffer | 25 mM Tris-HCl (pH 8.6), 192 mM glycine, and 0.1% SDS |
| SimplyBlue™ SafeStain | 80 mg L-1 Coomassie Brilliant Blue G-250, 35 mM HCl |
| 4× Gel loading buffer | 200 mM Tris-HCl (pH 6.8), 8% (w/v) sodium dodecyl sulfate, 0.4% (w/v) bromophenol blue, and 100 mM DTT |

### 8.5.2 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate and visualise DNA fragments. Agarose gels were prepared by dissolving 1% (w/v) agarose in 1× Tris-acetate-EDTA (TAE) buffer, followed by heating and adding SYBR® Safe DNA Gel Stain to a final concentration of 1×. Once the gel was poured and set, the agarose gel cassette was set up in a mini electrophoresis tank, and submerged in 1× TAE buffer. DNA samples were mixed with 6× DNA loading buffer to a final concentration of 1× and loaded into the solidified agarose gel. HyperLadder™ 1 kb (Bioline)

was run alongside the sample DNA fragments. The gel was subjected to electrophoresis for 40 min at 120 V. DNA bands within the gel were visualised by UV transillumination.

### 8.5.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins and assess the level of purity after different chromatography techniques. SDS-PAGE was performed using Bolt™ 4–12% Bis-Tris Plus precast protein gels in 1× Bolt™ 2-(N-morpholino)ethanesulfonic acid SDS running buffer. Protein samples were prepared by mixing with 4× gel loading buffer to a final concentration of 1×, heated at 80°C for 5 min and loaded into the precast protein gel alongside Novex™ Sharp Prestained Protein Standard. Electrophoresis was performed at 200 V for 22 min and gels were stained using SimplyBlue™ SafeStain. Gels were visualised and imaged using a Bio-5000 Plus digital imaging system (Microtek, Hsinchu, Taiwan).

## 8.6 Protein biochemistry

### 8.6.1 Recombinant protein expression

#### *8.6.1.1 Expression of adenyltransferase/IMP cyclohydrolase (ATIC) in pET28*

Cultures for protein expression were routinely grown to $OD_{600}$ 0.4–0.6 in LB medium, then cultures were cooled to an appropriate temperature for expression of the relevant construct. Protein expression was induced by adding isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. To determine whether titrating IPTG could aid in producing soluble ATIC, IPTG concentrations of 0.1 mM, 0.2 mM, 0.4 mM, 0.6 mM and 0.8 mM were used to induce protein expression. After induction, the temperature at the protein was expressed was also altered for each IPTG concentration. Expression temperatures included 37 °C for 3 h and 18°C and 26°C overnight.

#### *8.6.1.2 Standard protein expression conditions*

The proteins expressed in the pOPIN plasmids along with the coupling enzyme, dihydrodipicolinate reductase, required for the DHDPS kinetic assay (pET151, kindly provided

by Dr. Grant Pearce) were cultured and expressed by the method outlined here. Cultures for protein expression were routinely grown to an $OD_{600}$ of 0.4–0.6 in LB medium then cultures were cooled to an appropriate temperature for expression of the relevant construct. Protein expression was induced by adding IPTG to a final concentration of 1 mM.

### 8.6.1.3 *His$_6$TF-CLsoATIC aggregation prevention trials*

To prevent aggregation of His$_6$TF-*C*LsoATIC, additives were included in cultures prior to induction (Table 8.9). Buffers trialled for His$_6$TF-*C*LsoATIC disaggregation during protein purification are listed in Table 8.10.

**Table 8.9:** Composition of additives included in the culture of His$_6$TF-*C*LsoATIC.

| Additive | Concentration |
|---|---|
| Sorbitol | 0.1 M |
| | 0.5 M |
| | 1.0 M |
| NaCl | 0.1 M |
| | 0.5 M |
| | 1 M |
| Glycerol | 0.1% |
| | 0.5% |
| Benzyl alcohol | 0.1% |
| | 0.5% |
| Ethylene glycol | 0.1% |
| | 0.5% |

**Table 8.10:** Buffer solutions trialled for the prevention of $His_6$TF-*C*LsoATIC aggregation. Each buffer contained 0.1 mM tris(2-carboxyethyl)phosphine (TCEP).

| Buffer[1] | pH | NaCl Concentration (M) | Additives |
|---|---|---|---|
| Tris | 9.0 | 0.15 | |
| | 8.5 | 0.15 | |
| | 8.0 | 0.1 | |
| | | 0.15 | 0.15 M KCl |
| | | | 0.1 M arginine |
| | | | 0.1 M proline |
| | | | 0.1 M arginine, 0.1 M proline |
| | | | 0.1 % Triton-X |
| | | | 0.1 % n-dodecyl beta-d-maltoside |
| | | 0.2 | |
| | | 0.3 | |
| | | 0.4 | |
| | | 0.5 | |
| | | 1.0 | |
| | | 1.5 | |
| | 7.5 | 0.15 | |
| HEPES | 7.0 | 0.15 | |
| Bis-Tris | 6.5 | 0.15 | |
| | 6.0 | 0.15 | |
| MES | 5.5 | 0.15 | |

[1]Buffer concentrations were 50 mM

*8.6.1.4 Chaperone co-expression*

Cultures for protein expression were routinely incubated at 37°C with shaking (220 rpm) to $OD_{600}$ 0.2, at which point chaperone expression was induced with 0.5 mg/mL L-arabinose. The cultures were then grown to an $OD_{600}$ of 0.5 and expression of the protein of interest was induced with 1 mM IPTG. The cultures were incubated at 16°C for 18 h with shaking (220 rpm).

### 8.6.2 Harvesting of cells

Following protein expression, large scale cultures were harvested by centrifugation at 6,250 × *g*, for 20 min at 4°C. Cell pellets were then snap-frozen in liquid nitrogen and stored at -80°C.

### 8.6.3 Cell lysis

All cell lysis and purification was performed either on ice or at 4°C. Cell pellets were thawed and resuspended in ~3 mL/g of cell pellet (wet weight) in 1× PBS solution or the first buffer to be used in purification. Lysozyme was added to the cell resuspension at a final concentration of 0.5 mg/mL and incubated on ice for 30 min. Subsequently, 5 µg/mL of DNase I was added to the cell resuspension and incubated on ice for a further 15 min. Cells were lysed by sonication on ice using a UP200S Ultrasonic Processor (Hielscher, Teltow, Germany) at 70% amplitude, 0.5 s on and 0.5 s off for 10 min. Unlysed cells and debris were removed by centrifugation at 24,000 × *g*, for 25 min at 4°C. Supernatant was collected and stored at 4°C.

## 8.7 Protein purification

Various chromatographic techniques performed in this work used ÄKTA protein purification systems (GE Healthcare, Little Chalfont, United Kingdom) and gravity flow purification methods. All protein purification procedures were conducted at 4°C. Prior to any procedure, the columns or resin used for gravity flow techniques were pre-equilibrated with at least 3 column volumes (CV) of the initial purification buffer (Buffer A). In all cases, the cell free supernatant was loaded on to the appropriate column. To remove any unbound protein, the column was washed with an additional 3 CV of the initial purification buffer until stable baselines at 280 nm were achieved.

For immobilised metal affinity chromatography (IMAC), anion exchange chromatography and hydrophobic interaction chromatography, the protein was eluted with the respective Buffer A containing either a higher concentration of imidazole or sodium chloride, or a lower concentration of ammonium sulphate, respectively, using an isocratic or gradient elution profile. The eluate was fractionated and collected in 96-well plates. Following these

procedures, fractions were analysed by SDS-PAGE then fractions containing the protein of interest were conservatively selected and pooled with respect to their purity. After the final purification step, all proteins were aliquoted into ice-cold 1.7-mL micro-centrifuge tubes and snap-frozen in liquid nitrogen before long term storage at -80°C.

### 8.7.1 Purification of *C*Lso adenyltransferase/IMP cyclohydrolase

#### *8.7.1.1 ATIC with no fusion tag*

Anion exchange chromatography was employed as the first step for purifying adenyltransferase/IMP cyclohydrolase. The components of each buffer used for anion exchange chromatography are described in Table 8.11. The cell-free supernatant was applied to a 20-mL anion exchange column with a positively charged Q sepharose resin (GE Healthcare) , which had been pre-equilibrated with anion exchange buffer A. Once the cell free supernatant was loaded onto the column, it was washed with 3 CV of anion exchange buffer A. Bound proteins were eluted with a continuous gradient of anion exchange buffer B, up to 100% over 3 CV. The eluate was collected and fractions displaying peaks in the 280 nm absorbance trace were analysed for the protein of interest by SDS-PAGE before being pooled.

**Table 8.11:** Buffer solutions used for anion exchange chromatography.

| Buffer | Components |
|---|---|
| Anion exchange buffer A | 50 mM Tris (pH 8.5) |
| Anion exchange buffer B | 50 Mm Tris (pH 8.5), 1.5 M NaCl |

Hydrophobic interaction chromatography was used as a second purification step for adenyltransferase/IMP cyclohydrolase. The components of each buffer used for hydrophobic interaction chromatography are described in Table 8.12. Ammonium sulphate was added to the partially purified protein pooled from the anion exchange chromatography step to a final concentration of 1 M. The protein was then applied to a phenyl sepharose 20-mL column (GE Healthcare) pre-equilibrated with hydrophobic interaction buffer A and washed with 3 CV of this buffer. Bound proteins were eluted with a continuous gradient of hydrophobic interaction buffer B, up to 100% over 3 CV. Eluate fractions displaying any absorbance peaks at 280 nm were analysed by SDS-PAGE for the protein of interest prior to being pooled. The

protein was then further purified by size exclusion chromatography (outlined in Section 8.7.5).

**Table 8.12:** Buffer solutions used for hydrophobic interaction chromatography

| Buffer | Components |
|---|---|
| Hydrophobic interaction buffer A | 50 mM Tris (pH 8.5), 1 M ammonium sulphate |
| Hydrophobic interaction buffer B | 50 mM Tris (pH 8.5) |

### 8.7.2 Purification of ATIC-$His_6$ with chaperone co-expression

Purification of ATIC-$His_6$ co-expressed with chaperones used the same anion exchange protocol outlined in Section 8.7.1.1. After the first anion exchange chromatography step, the partially purified pooled protein was exchanged into anion exchange buffer A (Table 8.11) devoid of any sodium chloride using a HiPrep 26/10 desalting column (GE Healthcare). Fractions corresponding to the protein were pooled.

A second anion exchange process for the purification of the protein was performed using the same anion exchange buffers described in Table 8.11; however, a Q Sepharose High Performance column was used (GE Healthcare). The protein was then applied to size exclusion chromatography as the final purification step (outlined in Section 8.7.5).

### 8.7.3 Immobilised metal affinity chromatography

IMAC using a HisTrap FF 5 mL column (GE Healthcare) was employed as the first step to purify chaperone co-expressed dehydrorhamnose epimerase, dehydrorhamnose reductase, pyruvate kinase (PK), dihydrodipicolinate synthase (DHDPS), $His_6$TF-*C*LsoATIC and dihydrodipicolinate reductase. The column was pre-equilibrated with Buffer A for each protein and cell-free supernatant was applied to the column. The column was then washed with 5 CV of Buffer A, as used for pre-equilibration. Proteins were eluted from the column using an increasing gradient of elution Buffer B (Table 8.13). Fractions containing any absorbance peaks at 280 nm were analysed by SDS-PAGE for the protein of interest prior to being pooled.

**Table 8.13:** IMAC buffers. Buffer components are listed for each protein purified by IMAC in this thesis. Buffer A refers to the initial buffer used for column pre-equilibration. Buffer B refers to the elution buffer for each respective protein.

| Protein | Buffer Components | |
|---|---|---|
| | **Buffer A** | **Buffer B** |
| Pyruvate kinase<br>Dihydrodipicolinate synthase<br>Dehydrorhamnose epimerase<br>Dehydrorhamnose reductase | 1 x PBS (pH 7.4), 1 mM DTT, 5% glycerol and 40 mM imidazole | 1× PBS (pH 7.4), 1 mM DTT, 5% glycerol and 400 mM imidazole |
| $His_6$TF-*C*LsoATIC | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl and 40 mM imidazole | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl and 400 mM imidazole |
| | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl, 8 M Urea and 40 mM imidazole | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl, 8 M Urea and 400 mM imidazole |
| | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl, 6 M guanidine hydrochloride and 40 mM imidazole | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl, 6 M guanidine hydrochloride and 400 mM imidazole |
| Dihydrodipicolinate reductase | 50 mM $NaH_2PO_4$ (pH 8), 100 mM NaCl, 50 mM imidazole | 50 mM $NaH_2PO_4$ (pH 8), 100 mM NaCl, 50 mM imidazole |

### 8.7.4 Refolding of $His_6$TF-*C*LsoATIC

After IMAC purification of $His_6$TF-*C*LsoATIC under denaturing conditions (either 8 M urea or 6 M guanidine hydrochloride), the pooled protein was dialysed against 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl in order to remove the denaturant. The eluted protein was loaded into dialysis tubing (3.5K molecular weight cut-off) then incubated in 4 L of the destination buffer with stirring. The buffer was renewed every 2 h over an 8-h period. After dialysis, the pooled protein was purified by size exclusion chromatography (outline in Section 8.7.5).

### 8.7.5 Size exclusion chromatography

Size exclusion chromatography was employed as a final polishing step for all purification procedures described above using either a HiLoad Superdex 200 16/60 120 mL column or a Superdex 200 Increase 10/300 GL 24 mL column (GE Healthcare). All buffers used for size

exclusion chromatography throughout this thesis are listed in Table 8.14. Prior to being applied to the column, pooled protein from the previous purification step was concentrated to approximately 1% of the column volume. Eluted fractions corresponding to the 280 nm absorbance peak were analysed by SDS-PAGE for the protein of interest prior to being pooled.

**Table 8.14:** Size exclusion chromatography buffers used for each protein purified

| Expression method | Protein | Size exclusion buffer components |
|---|---|---|
| | ATIC with no fusion tag | 50 mM Tris (pH 8.5), 150 mM NaCl, 0.1 mM TCEP[1] |
| | Dihydrodipicolinate reductase | 20 mM Tris (pH 8.0) |
| Chaperone-coexpression | ATIC-$His_6$ | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl |
| | Pyruvate kinase<br>Dihydrodipicolinate synthase<br>Dehydrorhamnose epimerase<br>Dehydrorhamnose reductase | 1× PBS (pH 7.4), 1 mM DTT |
| pOPIN | $His_6$TF-*C*LsoATIC | 50 mM Tris (pH 8.5), 100 mM NaCl, 100 mM KCl |

## 8.8 Biophysical techniques

### 8.8.1 Circular dichroism spectroscopy

A Jasco J815 circular dichroism (CD) spectrophotometer (Jasco, Easton, MD) was used to record the CD spectra of purified proteins. Wavelength scans between 180 nm and 260 nm were performed on protein samples in an appropriate buffer. Data were collected in a 2-mm path length quartz cuvette at 20°C in 0.5-nm wavelength increments. Buffer data was subtracted from sample data, and data exceeding 600 V was removed. Data was converted to mean residue ellipticity by multiplying by the conversion factor (Equations 6.1 and 6.2). Resulting data was analysed using the DICHROWEB software package (Whitmore and Wallace 2008).

mean residue concentration = (number of amino acids in protein/molecular weight g/mol) × [sample] (mg/mL) [Equation 6.1]

conversion factor = 1/mean residue concentration × path length (mm) [Equation 6.2]

### 8.8.2 Differential scanning fluorimetry

A QuantStudio 3 real-time PCR system (ThermoFisher Scientific) was used to record the thermal stability of the purified protein with and without potential substrates. Reactions consisted of 0.5–1 mg/mL (15 µM) protein and 5× SYPRO® Orange dye (prepared as a 50× stock). Samples were prepared and measured in triplicate in 96-well thick wall plates. Fluorescence was measured for 10 s in 0.5 °C increments from 20−100°C.

### 8.8.3 Analytical ultracentrifugation

Sedimentation velocity experiments using analytical ultracentrifugation were performed in an XL-I Analytical Ultracentrifuge (Beckman Coulter, Pasadena, CA). Depending on the protein, the absorbance optical system was employed at wavelengths in the range of 280–285 nm. Initial scans were performed at 3,000 rpm to determine the optimal radial settings and wavelength. For these experiments, 400 µL of reference and 380 µL of sample solutions were loaded into 12-mm double sector cells with quartz or sapphire windows, and then mounted in an An-60 Ti four- or eight-hole rotor. Proteins were centrifuged at a rotor speed of 50,000 rpm at either 12°C or 20°C. Radial absorbance data were collected at a single wavelength without averaging, using a 0.003-cm step size for 80–120 scans. Solvent density, solvent viscosity and estimates of the partial specific volume for proteins were computed using the program SEDNTERP (Laue et al. 1992). Data were fitted to a continuous size distribution [c(s)] model using the program SEDFIT (Schuck 2000). This data was then fitted to a continuous mass distribution [c(M)] in SEDFIT to calculate the apparent molecular mass of the species observed in the [c(s)] distribution.

### 8.8.4 Small angle X-ray scattering

#### *8.8.4.1 Small angle X-ray scattering measurements*

Small angle X-ray scattering (SAXS) data was collected on the SAXS/WAXS beamline equipped with a Pilatus 1M detector (170 mm × 170 mm, effective pixel size, 172 µm × 172 µm) at the Australian Synchrotron. A sample detector distance of 1600 mm was used, providing a q range of 0.006–0.5 $Å^{-1}$. The purified protein was injected onto an inline Superdex S200 Increase 5/150 GL SEC column (GE Healthcare), equilibrated with 1× PBS buffer (pH 7.4) and

supplemented with 0.1% (w/v) sodium azide, at a flow rate of 0.35 mL/min$^{-1}$. Co-flow SAXS was used to minimise sample dilution and maximise signal to noise ratio. Scattering data was collected in 1-s exposures (λ = 1.0332 Å) over 500 frames, using a 1.5-mm glass capillary, at 12°C. Analysis of the scattering data was performed using the ATSAS software package (version 2.8.4). The molecular mass of the samples was estimated using the SAXS-MoW2 package.

*8.8.4.2 Small angle X-ray scattering data analysis*

All subsequent SAXS analyses were performed using the ATSAS software package (version 2.3) (Konarev et al. 2006). 2D intensity plots were radially averaged, normalised to sample transmission, and background subtracted using CHROMIXS. Guinier plots were analysed using PRIMUS (Konarev et al. 2003) to assess data quality. Indirect fourier transform of the data was performed using GNOM (Svergun 1992) to generate the *P(r)* distribution. Theoretical scattering curves were generated from atomic coordinates and compared with experimental scattering curves using CRYSOL (Svergun et al. 1995).

## 8.9 Enzyme kinetics

### 8.9.1 Steady state kinetics for adenyltransferase/IMP cyclohydrolase

Kinetic analysis for adenyltransferase/IMP cyclohydrolase was performed as described by Bulock et al. (2002). 5-formylaminoimidazole-4-carboxamide ribonucleotide (AICAR) and 10-formyltetrahydrofolate (FAICAR) were synthesised by Dr. Zoe Wright of the Wolan Laboratory (The Scripps Research Institute, San Diego, CA). All reactions were conducted in a 500-µL reaction volume in 1-cm pathlength quartz cuvettes at room temperature. Kinetic assays were carried out using a Cary 100 Bio UV/Vis spectrophotometer (Agilent Technologies, Santa Clara, CA).

The AICARFT assay reaction mixtures contained final concentrations of 66 mM Tris-HCl (pH 7.4), 50 mM potassium chloride, 100 µM AICAR (Sigma-Aldrich) and 300 µM 10-formyltetrahydrofolate. The reaction was initiated by the addition of *C*Lso ATIC (0.3 mg/mL) and monitored by the formation of tetrahydrofolate at 298 nm (Bulock et al. 2002).

For the reverse AICARFT assay, buffer concentrations were the same as the forward reaction. 100 µM of FAICAR and 100 µM of tetrahydrofolate were added to the reaction mixture and the reaction was monitored by the decrease in absorbance at 298 nm, which correlated with the consumption of tetrahydrofolate.

For the inosine monophosphate cyclohydrolase assay, reaction mixtures contained final concentrations of 100 mM Tris-Cl (pH 7.5), and 100 µM FAICAR. *C*Lso ATIC was then added to a final concentration of 100 nM to initiate the reaction. The reaction was monitored by the appearance of inosine monophosphate as an increase in absorbance at 248 nm.

### 8.9.2 Steady state kinetics for dihydrodipicolinate synthase

The $K_m$ values for *C*Lso DHDPS were unknown; therefore, apparent $K_m$ values were initially determined using a range of substrate concentrations. True kinetic constants ($K_m$ and $k_{cat}$) were then determined using a range concentrations approximately 10-fold above and below these values. Initial velocities were usually reproducible within 10%.

DHDPS activity was measured using a dihydrodipicolinate reductase coupled assay, as previously described (Yugari and Gilvarg 1965; Dobson et al. 2004). Stock solutions of (*S*)-ASA, pyruvate, (*S*)-lysine, NADPH and NADH were prepared fresh for each experiment. Approximately 20 µg of dihydrodipicolinate reductase was used in each assay to ensure it was present in excess. Assays were performed in 10-mm path-length, 1.5-mL reduced-volume acrylic cuvettes, in triplicate. The reaction was initiated by adding DHDPS, after the cuvette had pre-incubated at 30°C for 10 min. Activity was determined from the rate of absorbance decrease at 340 nm at 30°C. Table 8.15 shows the experimental design to determine the kinetic constants. Experiments were performed in either duplicate or triplicate.

**Table 8.15:** Parameters for the DHDPS coupled assay

| | [Initial] mM | Volume (µL) | [Final] mM |
|---|---|---|---|
| HEPES (pH 8) | 200 | 500 | 100 |
| NADPH | 5.4 | 30 | 0.162 |
| DHDPR | (in excess) | 100 | |
| *C*Lso DHDPS | 0.2 | 12.5 | 0.0025 |
| (*S*)-ASA | 0.04–2.5 | 50 | 0.002–0.128 mM |
| Pyruvate | 0.04–1.6 | 50 | 0.002–0.08 mM |
| $ddH_20$ | | 257.5 | |
| *Total* | | 1000 | |

To evaluate the effect of lysine on the catalytic potency of DHDPS, the kinetic experiments were designed as outlined in Table 8.16.

**Table 8.16:** Parameters for the DHDPS coupled assay with the allosteric inhibitor, lysine.

| | [Initial] mM | Volume (µL) | [Final] mM |
|---|---|---|---|
| HEPES (pH 8) | 200 | 500 | 100 |
| NADPH | 5.4 | 30 | 0.162 |
| DHDPR | (in excess) | 100 | |
| *C*Lso DHDPS | 0.2 | 12.5 | 0.0025 |
| (*S*)-ASA | 0.04–2.5 | 50 | 0.002–0.128 |
| Pyruvate | 0.04–1.6 | 50 | 0.002–0.08 |
| lysine | 1, 2, 4, 8, 16 | 50 | 0.05, 0.1, 0.2, 0.4, 0.8 |
| $ddH_20$ | | 207.5 | |
| Total | | 1000 | |

All data was analysed using GraphPad Prism. Data were fitted to the appropriate models as judged by the $R^2$ value.

### 8.9.3 Assay for pyruvate kinase activity

The enzyme assay was modified from that described by Bucher and Pfleiderer (1955) and recently adapted in Donovan et al. (2016). An aliquot of purified *C*Lso PK (0.4 mg/mL) was added to a solution containing 10 mM HEPES (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 2 mM phosphoenolpyruvate, 2 mM ADP, 2 mM fructose-1,6-bisphosphate, 0.12 mM NADH and 2.2

units of lactate dehydrogenase. Activity was determined from the rate of absorbance decrease at 340 nm at 25°C.

## 8.10 X-ray crystallography

### 8.10.1 Crystallisation

Crystallisation studies were performed using freshly purified protein, concentrated to 10–20 mg/mL with an appropriate molecular weight cut off spin concentrator. Initial protein crystallisation trials were performed in-house. All trays were inspected every 2–3 days for crystal growth. Initial crystallisation trials of *C*Lso DHDPS and PK used the PACT Suite, JCSG+ Suite, and Morpheus crystal screens (Newman et al. 2005; Newman et al. 2008). These trials were conducted using the sitting drop vapour diffusion method, at 20°C, with droplets consisting of 150 nL of protein solution and 150 nL of reservoir solution.

### 8.10.2 Data collection, processing and structure refinement

*8.10.2.1 Data collection*

For X-ray data collection, crystals were mounted onto cryo-loops of an appropriate size for the crystal of interest and cryo-protected by soaking in 85% reservoir solution and 15% glycerol-ethylene glycol (50:50 mix) prior to being flash frozen in liquid nitrogen. Crystals that had been flash cooled were mounted onto the beamline in a cold nitrogen stream at -173°C. Diffraction data was collected on the MX2 beamline at the Australian Synchrotron with the EIGER × 16M detector (λ = 0.9537 Å).

*8.10.2.2 Data processing*

Data was processed and scaled using XDS and AIMLESS from the CCP4i2 program suite (Potterton et al. 2018). The resulting intensity data were analysed using PHENIX XTRIAGE and the PHENIX REFLECTION FILE EDITOR (Adams et al. 2010) was used to generate a new $R_{free}$ set in thin resolution shells. The output allowed the true resolution of the data, the signal to noise ratio, multiplicity and completeness of the data to be assessed. An estimate of the number of

molecules in the asymmetric unit was obtained using the MATTHEWS COEFFICIENT program (Matthews 1968) from the CCP4 suite.

*8.10.2.3 Model building and structure refinement*

The coordinate file of the search model was prepared using CHAINSAW (Stein 2008) from the CCP4 Suite, which uses a sequence alignment of the search model protein and the target protein to edit the coordinate file. Molecular replacement was used to estimate the phases from a search model using the CCP4 program PHASER (McCoy et al. 2007). Structure refinement was initially performed using REFMAC (Murshudov et al. 2011) in CCP4 and continued thereafter in PHENIX REFINE (Adams et al. 2010). Simulated annealing was performed to eliminate model bias the first time that refinement was conducted in PHENIX REFINE. Iterative improvement of both the map and the model was performed using alternate cycles of refinement and residue-by-residue analysis in COOT (Emsley and Cowtan 2004). Some residues were manually repositioned in COOT to better fit the electron density. Water molecules were added using the "add waters" function in COOT, and manually assessed for fit to the observed spherical electron density. Ligands were added using COOT and manually modelled into the observed electron density map. Coordinate and restraint files for ligands not in the COOT library were generated via SKETCHER in the CCP4 suite. Following each cycle of refinement, the output was monitored for a decrease in both the residual factor ($R_{factor}$) and the free $R_{factor}$ ($R_{free}$). All final images were generated in PyMOL.

## 8.11 Gene expression analysis methods

### 8.11.1 Primer design

Primers were designed using the National Centre for Biotechnology Information (NCBI) Primer-Basic Local Alignment Search Tool (Primer-BLAST) (Ye et al. 2012), which uses Primer 3 (Rozen et al. 200) to design PCR primers, and BLAST (Altschul et al. 1990) to screen primers against a user-selected database in order to avoid primer pairs that can cause non-specific amplification. The Primer-BLAST search was limited to a PCR product size 100–110 bp. The genes and the primers used to amplify the gene products are described in Table 8.17.

**Table 8.17:** Genes and primer sequences used in gene expression analysis

| Gene Role | Gene | Accession Number | Protein | Primer Sequence |
|---|---|---|---|---|
| Target | *dapA* | KJZ81861.1 | Dihydrodipicolinate synthase | 5′-CACGGAGGTGTGGGTTGTAT-3′<br>3′-GTGCTTGACGATAATCCCCCT-5′ |
| | *secDF* | KJZ81761.1 | Protein-export membrane protein SECDF | 5′-GCCTCCTCGTAAATCCAGCC-3′<br>3′-GGAATGCTTGTTGCATTGCCTA-5′ |
| | *znMC* | WP_034442221.1 | Serralysin | 5′-CGTGTATACCGATCATCGCCT-3′<br>3′-CACCACTTTGCTCCATTCGC-5′ |
| | *purH* | KJZ82278.1 | Adenyltransferase/IMP cyclohyrolase | 5′-TGCTTCTTATGATGCGGCGA-3′<br>3′-CCCCATAACGCATCTCTTGC-5′ |
| | *rmlC* | KJZ80689.1 | Dehydrorhamnose epimerase | 5′-TTTTCAGAAACCGCCGCATG-3′<br>3′-ATAGGTCGGGCTATCTTTGCG-5′ |
| | *argJ* | KJZ82379.1 | Glutamate N-acetyltransferase | 5′-GGACAAATTTGCCGCATGGT-3′<br>3′-TGCGCAATAAAATGCACAGCA-5′ |
| | *rmlD* | KJZ81768.1 | Dehydrorhamnose reductase | 5′-GCGTGCCGTGCATTCATATT-3′<br>3′-CCATAAATGCCCAGCGGGTT-5′ |
| | *coaBC* | KJZ81994.1 | Phosphopantothenoylcysteine decarboxylase/ phosphopantothenate synthase | 5′-CCACGCCGCAACCATTATTT-3′<br>3′-TCAGCTACGCAACGTAATGTC-5′ |
| | *pykF* | KJZ82234.1 | Pyruvate kinase | 5′-CGTGTTTTTCTTCCGCATGC-3′<br>3′-CCTGTCCCCTTCACGCATAG-5′ |
| Housekeeping | *recA* | KJZ80672 | Recombinase A | 5′-TACGCCCTTTGGGAAAACCA-3′<br>3′-TGCAGCTTGGCTCTCAAAGT-5′ |
| | *gapA* | KJZ82286.1 | NAD-dependent glyceraldehyde-3-phosphate dehydrogenase | 5′-GTTCGTCCGTTCGAGTTCCT-3′<br>3′-GCAAAATCTCTCATCGCCGC-5′ |
| | *rplD* | KJZ81844.1 | Large subunit ribosomal protein | 5′-GTCAAAGGGTCGTTCCGAGA-3′<br>3′-CCACGAAATTGAGGGGCAGA-5′ |
| | *rpb* | KJZ81365.1 | DNA-directed RNA polymerase beta subunit | 5′-CCACAAGATACAATCGCCGC-3′<br>3′-GCGGGAAAAGGTTTACTGCG-5′ |

### 8.11.2 DNA extraction

Adult tomato potato psyllids were collected from *C*Lso-infected and -uninfected colonies. DNA was extracted from five psyllids from each colony using cetyl trimethylammonium bromide (CTAB) extraction as described by Beard and Scott (2013). Tomato potato psyllids were processed as whole individuals. Psyllids were ground in a micro-centrifuge tube with a micropestle, then 100 µL of 2× CTAB buffer was added to the sample and incubated for 2 hours at 65°C. Chloroform:isoamylalcohol (24:1; 100 µL) was added to the psyllid samples and agitated for 2–5 min prior to centrifugation at 6,000 × *g* for 15 min. The supernatant was transferred to a fresh micro-centrifuge tube containing 80 µL of isopropanol and inverted. The sample was then incubated at -20°C overnight. Samples were centrifuged at 16,000 × *g* for 15 min. The supernatant was discarded and the pellet was washed with 70% ethanol. The pellet was air dried and resuspended in elution buffer.

### 8.11.3 Primer testing

A semi-nested qPCR protocol was used in order to determine the titre of *C*Lso present in DNA extracted from infected psyllids. *C*Lso DNA was amplified using SsoAdvanced™ universal SYBR Green Supermix and the qPCR reaction was performed in 10-µL reactions. For the reaction, 1 µL of template, 50 nM primer LsoF, 300 nM of primer Lso16SF and 300 nM of primer Lso16SRI were used. The qPCR parameters are presented in Table 8.2. The psyllid ITS2F (300 nM) and psyllid ITS2R (300 nM) primers were used as an internal qPCR control (Table 8.18). Samples were heated to 95°C for 3 min, followed by 40 cycles of 95°C for 20 s, 60°C for 20 s, and 72°C for 20 s, with data capture during the 72°C step, followed by a melt curve analysis (65–90°C). The qPCR primers for the gene expression analysis were tested using DNA extracted from psyllids and the primer concentration and amplification parameters described above. All samples were run in triplicate. The DNA products were run on a 1% agarose gel in order to determine primer specificity (see Appendix, Figure i).

**Table 8.18:** Primers used to determine *C*Lso titres in DNA samples and to serve as internal psyllid controls

| Primer | Sequence |
|---|---|
| *C*LsoF | 5′-GTCGAGCGCTTATTTTTAATAGGA-3′ |
| *C*Lso16SF | 5′-ATACCGTATACGCCCTGAGAAG-3′ |
| *C*Lso16SRI | 3′-CGTAGCCTTGGTAGGCATT-5′ |
| ITS2F | 5′-AAGCGACGTGTGGAAGAACC-3′ |
| ITS2R | 3′-GTTGTGTGTGTCCGGGGAAG-5′ |

### 8.11.4 RNA extraction and cDNA synthesis

Gene expression experiments were performed using tomato plants, *Solanum lycopersicum* L. Prior to RNA extraction, plant samples were stored in RNAlater. The Geneaid Total RNA Mini Kit and the Aurum total RNA fatty and fibrous tissue kit were used for RNA extraction from psyllid and plant samples, respectively, as described by the manufacturer's instructions. The isolated RNA was treated with DNase (Promega, Madison, WI) as described by the manufacturer. RNA was amplified using primers described in Table 8.18 by conventional PCR to test for the presence of any genomic DNA. cDNA was synthesized from 1 µg of total RNA using the iScript cDNA Synthesis Kit (BioRad) as described by the manufacturer. Initial concentrations of RNA were normalised for each sample.

### 8.11.5 Quantitative reverse transcription PCR (qRT-PCR)

All qRT-PCR reactions were performed on a StepOne Plus real time PCR system (Thermo Fisher Scientific). SsoAdvanced™ universal SYBR reagent was mixed with appropriate primers in a reaction volume of 20 µL, and added to a 96-well PCR plate. cDNA (5 µL) was added to the appropriate well. Samples were run in triplicate ensuring that infected and uninfected psyllid and plant samples were run for each primer pair, along with primers for the housekeeping gene. Recombinase A (*recA*) and the DNA-directed RNA polymerase beta subunit (*rpb*) were used as a reference gene to provide normalisation of relative quantification for the target genes upon *C*Lso infection (Appendix, Figure ii). Thermal cycling parameters were 95°C for 10 s, followed by 40 cycles of 60°C for 15 s with a melt-curve analysis from 65–95°C in 0.5°C increments every 2 s.

### 8.11.6 Data analysis

The relative quantification of gene expression level was determined by the comparative CT method $2^{-\Delta\Delta}C_T$ (Livak and Schmittgen 2001), where $^{\Delta\Delta}CT$ is calculated by the following equation:

$$^{\Delta\Delta}C_T = (C_{T,\ target}–C_{T,\ reference})CLso\ infected–(C_{T,\ target}–C_{T,\ reference})Control$$

Standard deviation and P value were calculated using a Student's *t* test in as described by Yuan et al. (2006).

## 8.12 References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., and McCoy, A.J. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66(2):213-221.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215(3):403-410.

Beard, S.S., and Scott, I.A. 2013. A rapid method for the detection and quantification of the vector-borne bacterium '*Candidatus* Liberibacter solanacearum'in the tomato potato psyllid, *Bactericera cockerelli*. Entomologia Experimentalis et Applicata 147(2):196-200.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic acids research 35(6):e45.

Bücher, T., and Pfleiderer, G. 1955. Pyruvate kinase from muscle. Pyruvate phosphokinase, pyruvic phosphoferase, phosphopyruvate transphosphorylase, phosphate-transferring enzyme II, etc. Phosphoenolpyruvate + ADP ⇌ pyruvate + ATP. Methods in Enzymology 1:435-440.

Bulock, K.G., Beardsley, G.P., and Anderson, K.S. 2002. The kinetic mechanism of the human bifunctional enzyme ATIC (5-amino-4-imidazolecarboxamide ribonucleotide transformylase/inosine 5'-monophosphate cyclohydrolase) a surprising lack of substrate channeling. Journal of Biological Chemistry 277(25):22168-74.

Dobson, R.C., Griffin, M.D., Roberts, S.J., and Gerrard, J.A. 2004. Dihydrodipicolinate synthase (DHDPS) from *Escherichia coli* displays partial mixed inhibition with respect to its first substrate, pyruvate. Biochimie 86(4-5):311-5.

Donovan, K.A., Zhu, S., Liuni, P., Peng, F., Kessans, S.A., Wilson, D.J., and Dobson, R.C. 2016. Conformational dynamics and allostery in pyruvate kinase. Journal of Biological Chemistry 29*1*(17):9244-9256.

Emsley, P., and Cowtan, K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallography D60:2126-2132.

Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. 2006. ATSAS 2.1, a program package for small-angle scattering data analysis. Journal of Applied Crystallography 39(2):277-286.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H., and Svergun, D.I. 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography 36(5):1277-1282.

Laue, T.M., and Stafford, W.F. 1992. Modern applications of analytical ultracentrifugation. Annual Review of Biophysics and Biomolecular Structure 28(1):75-100.

Livak, K.J., and Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods 25(4):402-408.

Matthews, B.W. 1968. Solvent content of protein crystals. Journal of Molecular Biology 33(2):491-497.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. 2007. Phaser crystallographic software. Journal of Applied Crystallography 40(4):658-674.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D: Biological Crystallography 67(4):355-367.

Newman, J., Egan, D., Walter, T.S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J.L., Stuart, D.I., and Perrakis, A. 2005. Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. Acta Crystallographica Section D: Biological Crystallography 61(10):1426-1431.

Newman, J., Pham, T.M., and Peat, T.S. 2008. Phoenito experiments: combining the strengths of commercial crystallization automation. Acta Crystallographica Section F: Structural Biology and Crystallization Communications 64(11):991-996.

Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P.R., Jenkins, H.T., Keegan, R., Krissinel, E., Stevenson, K., and Lebedev, A. 2018. CCP4i2: the new graphical user interface to the CCP4 program suite. Acta Crystallographica Section D: Structural Biology 74(2):68-84.

Rozen, S., and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols, Humana Press, Totowa, NJ.

Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophysical Journal 78(3):1606-1619.

Whitmore, L., and Wallace, B.A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers: Original Research on Biomolecules 89(5):392-400.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13(1):134.

Yuan, J.S., Reed, A., Chen, F., and Stewart, C.N. 2006. Statistical analysis of real-time PCR data. BMC Bioinformatics 7(1):85.

Yugari, Y., and Gilvarg, C. 1962. Coordinated end-product inhibition in lysine synthesis in *Escherichia coli*. Biochimica et biophysica acta 62(3):612-4.

# Appendix

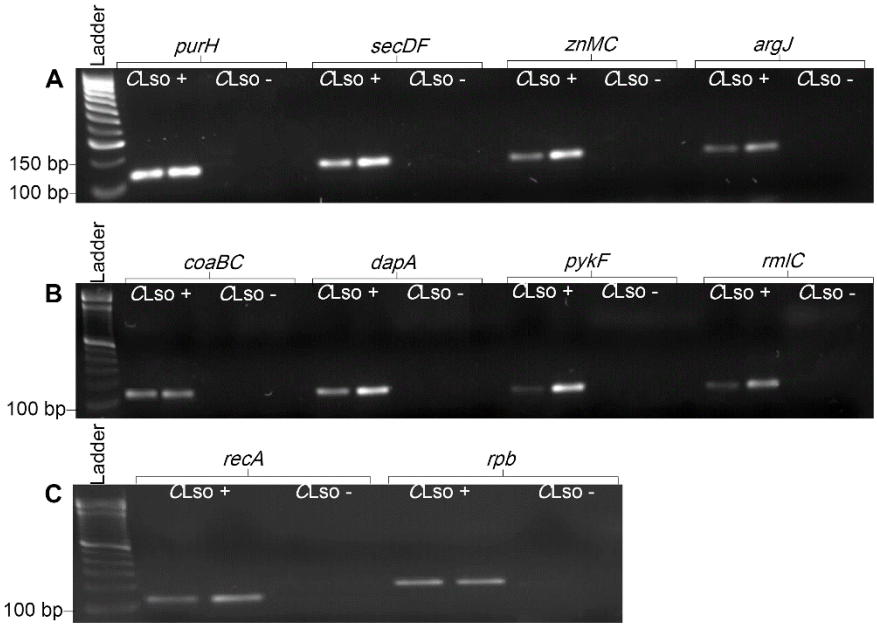



**Figure i:** Primer specificity of the qPCR primers used in the gene expression analysis. **A**, **B** and **C** show DNA extracted from *C*Lso infected and uninfected psyllids and amplified with the respective primers. Each primer shows specific amplification of *C*Lso genes and the absence of non-specific DNA amplificaton in *C*Lso uninfected psyllids.

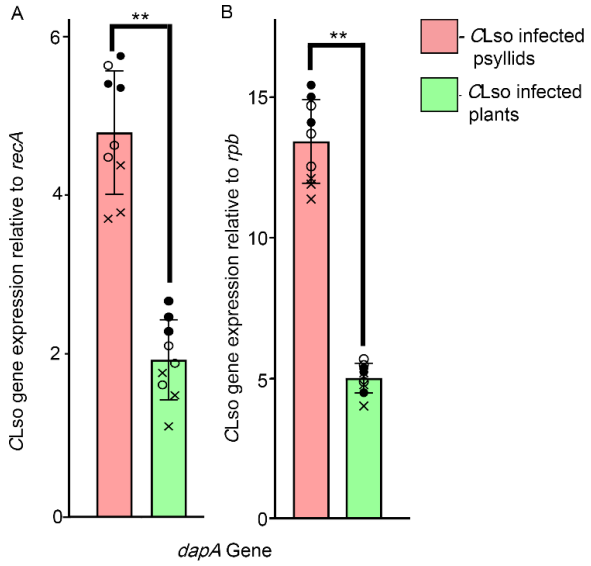



**Figure ii:** Gene expression of *C*Lso genes in psyllid and in planta based on qRT-PCR. A and B show the gene expression in relation to the housekeeping gene *recA* (recombinase A, KJZ80672, DJ66_1282) and *rpb* (DNA-directed RNA polymerase beta subunit (KJZ81364.1, DJ66_1258 ).Mean ± standard deviation for biological (o, ●, ×) and three technical replicates. Asterisks indicate statistically significant differences in expression levels between *C*Lso infected psyllids and plants (*P ≤ 0.05, **P ≤ 0.01; *t*-test).

## Sequences of proteins expressed in this study

### ATIC–no tag (expressed from pET28)

```
MDFSHRKYCG SDEILVKNAL ISVYDKTGIV EFSRRLLSWG IKLVSTGGTC QLLEKENIPV  60
TNISDITGFP EIMDGRVKTL HPKIYGGLLG IRDNPEHMMA MREHGLESID LVVVNLYPFE 120
ESCLREDEYH TIVENIDIGG PSMIRAAAKN HDYVTVLTDS KDYPLFLSEM ESNNGKVLLN 180
FRKKMALQAF FRTASYDAAI CRWFANAKNE DFPDYLNITG TKKQEMRYGE NPHQKAALYS 240
IFNKKSGIAN ATLVQGKPLS YNNINDLDAA YELVSEFSSE DCAACVIVKH TNPCGVAIAD 300
TLLEAYRRAL SCDPMSAFGG IVAFNQALDQ EVAREVIKIF TEVIVAPKIS KEAEVVISEK 360
PNMRFVKTFE LSDPRAQDIV FKKVSGGILV QTRDNNVVDD QELTVVTKRS PTPKELSDMK 420
FALKIVKHVK SNAVVFAKEG RTIGIGSGQT SRVDSTRFAA IKSQGISAQE GIKSVTSGAV 480
IASEAFYPFP DGILEAIKAG VTAVIQPGGS IRDLEIISVA DQHNIAMVFT GTRHFRH    537
```

### $His_6$–AICARFT domain of ATIC (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPD FSHRKYCGSD EILVKNALIS VYDKTGIVEF SRRLLSWGIK  60
LVSTGGTCQL LEKENIPVTN ISDITGFPEI MDGRVKTLHP KIYGGLLGIR DNPEHMMAMR 120
EHGLESIDLV VVNLYPFEES CLREDEYHTI VENIDIGGPS MIRAAAKNHD YVTVLTDSKD 180
YPLFLSEMES NNGKVLLNFR KKMALQAFFR TASYDAAICR                       220
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

### $His_6$–IMPCH domain of ATIC (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPK QEMRYGENPH QKAALYSIFN KKSGIANATL VQGKPLSYNN  60
INDLDAAYEL VSEFSSEDCA ACVIVKHTNP CGVAIADTLL EAYRRALSCD PMSAFGGIVA 120
FNQALDQEVA REVIKIFTEV IVAPKISKEA EVVISEKPNM RFVKTFELSD PRAQDIVFKK 180
VSGGILVQTR DNNVVDDQEL TVVTKRSPTP KELSDMKFAL KIVKHVKSNA VVFAKEGRTI 240
GIGSGQTSRV DSTRFAAIKS QGISAQEGIK SVTSGAVIAS EAFYPFPDGI LEAIKAGVTA 300
VIQPGGSIRD LEIISVADQH NIAMVFTGTR HFRH                             334
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

## $His_6$–ATIC (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPD FSHRKYCGSD EILVKNALIS VYDKTGIVEF SRRLLSWGIK  60
LVSTGGTCQL LEKENIPVTN ISDITGFPEI MDGRVKTLHP KIYGGLLGIR DNPEHMMAMR 120
EHGLESIDLV VVNLYPFEES CLREDEYHTI VENIDIGGPS MIRAAAKNHD YVTVLTDSKD 180
YPLFLSEMES NNGKVLLNFR KKMALQAFFR TASYDAAICR WFANAKNEDF PDYLNITGTK 240
KQEMRYGENP HQKAALYSIF NKKSGIANAT LVQGKPLSYN NINDLDAAYE LVSEFSSEDC 300
AACVIVKHTN PCGVAIADTL LEAYRRALSC DPMSAFGGIV AFNQALDQEV AREVIKIFTE 360
VIVAPKISKE AEVVISEKPN MRFVKTFELS DPRAQDIVFK KVSGGILVQT RDNNVVDDQE 420
LTVVTKRSPT PKELSDMKFA LKIVKHVKSN AVVFAKEGRT IGIGSGQTSR VDSTRFAAIK 480
SQGISAQEGI KSVTSGAVIA SEAFYPFPDG ILEAIKAGVT AVIQPGGSIR DLEIISVADQ 540
HNIAMVFTGT RHFRH                                                 555
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

## $His_6$–Trigger Factor–ATIC (expressed from pOPINTF)

```
MAHHHHHHQV SVETTQGLGR RVTITIAADS IETAVKSELV NVAKKVRIDG FRKGKVPMNI  60
VAQRYGASVR QDVLGDLMSR NFIDAIIKEK INPAGAPTYV PGEYKLGEDF TYSVEFEVYP 120
EVELQGLEAI EVEKPIVEVT DADVDGMLDT LRKQQATWKE KDGAVEAEDR VTIDFTGSVD 180
GEEFEGGKAS DFVLAMGQGR MIPGFEDGIK GHKAGEEFTI DVTFPEEYHA ENLKGKAAKF 240
AINLKKVEER ELPELTAEFI KRFGVEDGSV EGLRAEVRKN MERELKSAIR NRVKSQAIEG 300
LVKANDIDVP AALIDSEIDV LRRQAAQRFG GNEKQALELP RELFEEQAKR RVVVGLLLGE 360
VIRTNELKAD EERVKGLIEE MASAYEDPKE VIEFYSKNKE LMDNMRNVAL EEQAVEAVLA 420
KAKVTEKETT FNELMNQQAS SGLEVLFQGP DFSHRKYCGS DEILVKNALI SVYDKTGIVE 480
FSRRLLSWGI KLVSTGGTCQ LLEKENIPVT NISDITGFPE IMDGRVKTLH PKIYGGLLGI 540
RDNPEHMMAM REHGLESIDL VVVNLYPFEE SCLREDEYHT IVENIDIGGP SMIRAAAKNH 600
DYVTVLTDSK DYPLFLSEME SNNGKVLLNF RKKMALQAFF RTASYDAAIC RWFANAKNED 660
FPDYLNITGT KKQEMRYGEN PHQKAALYSI FNKKSGIANA TLVQGKPLSY NNINDLDAAY 720
ELVSEFSSED CAACVIVKHT NPCGVAIADT LLEAYRRALS CDPMSAFGGI VAFNQALDQE 780
VAREVIKIFT EVIVAPKISK EAEVVISEKP NMRFVKTFEL SDPRAQDIVF KKVSGGILVQ 840
TRDNNVVDDQ ELTVVTKRSP TPKELSDMKF ALKIVKHVKS NAVVFAKEGR TIGIGSGQTS 900
RVDSTRFAAI KSQGISAQEG IKSVTSGAVI ASEAFYPFPD GILEAIKAGV TAVIQPGGSI 960
RDLEIISVAD QHNIAMVFTG TRHFRH                                     986
```

Red indicates $His_6$ tag, blue indicates Trigger Factor fusion protein and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

## ATIC–$His_6$ (expressed from pET24)

```
MDFSHRKYCG SDEILVKNAL ISVYDKTGIV EFSRRLLSWG IKLVSTGGTC QLLEKENIPV  60
TNISDITGFP EIMDGRVKTL HPKIYGGLLG IRDNPEHMMA MREHGLESID LVVVNLYPFE 120
ESCLREDEYH TIVENIDIGG PSMIRAAAKN HDYVTVLTDS KDYPLFLSEM ESNNGKVLLN 180
FRKKMALQAF FRTASYDAAI CRWFANAKNE DFPDYLNITG TKKQEMRYGE NPHQKAALYS 240
IFNKKSGIAN ATLVQGKPLS YNNINDLDAA YELVSEFSSE DCAACVIVKH TNPCGVAIAD 300
TLLEAYRRAL SCDPMSAFGG IVAFNQALDQ EVAREVIKIF TEVIVAPKIS KEAEVVISEK 360
PNMRFVKTFE LSDPRAQDIV FKKVSGGILV QTRDNNVVDD QELTVVTKRS PTPKELSDMK 420
FALKIVKHVK SNAVVFAKEG RTIGIGSGQT SRVDSTRFAA IKSQGISAQE GIKSVTSGAV 480
IASEAFYPFP DGILEAIKAG VTAVIQPGGS IRDLEIISVA DQHNIAMVFT GTRHFRHHHH 540
HHH                                                             543
```

Red indicates $His_6$ tag.

## DHDPS–$His_6$ (expressed from pET24)

```
MFQRSIPALI TPFTKDNLID EDSFVDHIEW QISEGSSGLV PAGTTGESST LSYEEHCRVV  60
ELCVKTAAGR VPVMAGAGSN NTKESIELAQ YAQNTGADAL LVVVPYYNKP NKKGLLAHFG 120
SIANAVSLPI YIYNNPSRTV IEMDVDTMAE LVKTYSNIVG VKDATGRIEL ASGQRIACGS 180
DFIQLSGDDS SALGFNVHGG VGCISVTANV APRICAEFQK AISEGDYRQA LEYQDKLFPL 240
HQALFIEPSI SSVKYALSRL GRNVSLVVRA PMVSILEKET MFAIDQALDH IGLCAGHHHH 300
HH                                                              302
```

Red indicates $His_6$ tag.

## PK–$His_6$ (expressed from pET24)

```
MQSLRRIKII STLGPASFSE DLIRRLHEEG TDLFRINMSH TSHEKMRDLI RKIRVVESQS  60
QRPVGILIDL QGPKFRVGKF EGSKIYLKQG QNFTLDNKDS LGNVDRVFLP HAEVFDSVKV 120
GDRLLIDDGK VKLCVKGTGS DFIKCEVVSG SSISDRKGIS FPDTLLTTQA LTPKDREDLH 180
AALQTSEVDW VALSFIQSTE DLVEIRKIIS PQQIGLMSKI EKSQAIDYAT EIIKLSDAVM 240
VARGDLGVEM PLESIPGIQK KLIRIARQLG KPIVIATQML ESMISSPVPT RAEVSDVATA 300
VFEEADAIML SAETASGSYP VDAVKMMALV ASSAEKDPSW LDMRSLRKIE PNETGADVIS 360
SAARQIAETL RLSAIVCYTA SGQTGLRTAR ERPKLAIVAL SPVIRTARRL SLVWGVHCVV 420
TEDASGFDDM VNRACRIVVE QGFGKPGDRI IISAGLPLRT PGSTNMLRIV YIGEDGLSGV 480
HHHHHH                                                          486
```

Red indicates $His_6$ tag.

## Serralysin–$His_6$ (expressed from pET24)

```
MDDIKPVYTT ESIADYLLRD SYWANYFLKY QPTVYTDHRL SFDVTELNAN GQEIVRWAFR  60
EWSKVVNITF EETSNLSDIQ FKSNDFGYIC IPKNNFRYFM EINLDARDVD FYGIDKGTIL 120
SHNAIHEIGH ALGLLHAGPY NGGSPTYGVD NQYANDSFLT SVMSYFDQNE NRTTDASFGY 180
CATPMVADIV AVQKIYGAPK TLVGDTIHKI EVKQGYSPFI QTIYDSGGSD LLDVQSVGTF 240
DNYIDLNPES WSSIGGYKKN VTIARDTIIE SVIGGAGNDT IIGNSANNTL FKSGGNDTFY 300
GADGWDVMVY SMLKNDYKIY RFKNEAIIYD MFRNEVDDII DVESLRFSND QIDLADIKQQ 360
SILEYTASYG DLVQCIGNDV NASKKHFVDF GLREGRDITF NPYLYLGSYD DLRQVFGSDK 420
DSATRHYITH GFREGRNPGF MYRDDSMQHS FSLSEKSQEH KYTSTFFDPK EDLSENFFEW 480
PNSFQDTIFD FSEHHHHHH                                            499
```

Red indicates $His_6$ tag.

## RmlC–$His_6$ (expressed from pET24)

```
MEIINPVRIL KTKILEDDRG WFSETYNDHC FKKNGFKDVF VQDNHSFSLN TGTIRGLHFQ  60
KPPHAQAKIV RCVSGRIFDV AVDVRKDSPT YGNWISLEIS ASSGLQIYIP IGFAHGFMTL 120
EPNTEVIYKV TDFYSSSHDS GVLWQDKSIS IRWPLLEKIS PSLSDKDMNL PFLNQLDSPF 180
AYDGFPLLPF DMERDMSCVH HHHHH                                     205
```

Red indicates $His_6$ tag.

## RmlD–$His_6$ (expressed from pET24)

```
MKFLVIGNNG QVATSLFNLS GKNTEVIRLG RPDIDLLEPK DFTKIFLSFA PDVIINPAAY  60
TAVDKAEDEP HNAFAINTEG AGAVAAAAES IGVPCIHIST DYVFDGSSQT PICETSPTNP 120
LGIYGKSKLA GEKKVMSCTD NYVILRTAWV YSLFGTNFFL SMLRLSKEYR EISVVCDQFG 180
TPTSASQIAI AVIQIARNLV NNSDNSLRGI FHMTANGGPV SWADFAEYIF LMSTELGGNS 240
CKVRRISTNQ YPTKARRPAY SFLDCSKLEK IHNVRISSWK EGVRNVFIKH HHHHH      295
```

Red indicates $His_6$ tag.

## $His_6$–DHDPS (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPF QRSIPALITP FTKDNLIDED SFVDHIEWQI SEGSSGLVPA  60
GTTGESSTLS YEEHCRVVEL CVKTAAGRVP VMAGAGSNNT KESIELAQYA QNTGADALLV 120
VVPYYNKPNK KGLLAHFGSI ANAVSLPIYI YNNPSRTVIE MDVDTMAELV KTYSNIVGVK 180
DATGRIELAS GQRIACGSDF IQLSGDDSSA LGFNVHGGVG CISVTANVAP RICAEFQKAI 240
SEGDYRQALE YQDKLFPLHQ ALFIEPSISS VKYALSRLGR NVSLVVRAPM VSILEKETMF 300
AIDQALDHIG LCAG                                                  321
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

## $His_6$–PK (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPQ SLRRIKIIST LGPASFSEDL IRRLHEEGTD LFRINMSHTS  60
HEKMRDLIRK IRVVESQSQR PVGILIDLQG PKFRVGKFEG SKIYLKQGQN FTLDNKDSLG 120
NVDRVFLPHA EVFDSVKVGD RLLIDDGKVK LCVKGTGSDF IKCEVVSGSS ISDRKGISFP 180
DTLLTTQALT PKDREDLHAA LQTSEVDWVA LSFIQSTEDL VEIRKIISPQ QIGLMSKIEK 240
SQAIDYATEI IKLSDAVMVA RGDLGVEMPL ESIPGIQKKL IRIARQLGKP IVIATQMLES 300
MISSPVPTRA EVSDVATAVF EEADAIMLSA ETASGSYPVD AVKMMALVAS SAEKDPSWLD 360
MRSLRKIEPN ETGADVISSA ARQIAETLRL SAIVCYTASG QTGLRTARER PKLAIVALSP 420
VIRTARRLSL VWGVHCVVTE DASGFDDMVN RACRIVVEQG FGKPGDRIII SAGLPLRTPG 480
STNMLRIVYI GEDGLSGVKL SRPV                                        504
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

## $His_6$–RmlC (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPE IINPVRILKT KILEDDRGWF SETYNDHCFK KNGFKDVFVQ  60
DNHSFSLNTG TIRGLHFQKP PHAQAKIVRC VSGRIFDVAV DVRKDSPTYG NWISLEISAS 120
SGLQIYIPIG FAHGFMTLEP NTEVIYKVTD FYSSSHDSGV LWQDKSISIR WPLLEKISPS 180
LSDKDMNLPF LNQLDSPFAY DGFPLLPFDM ERDMSCV                          217
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

## $His_6$–RmlD (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPK FLVIGNNGQV ATSLFNLSGK NTEVIRLGRP DIDLLEPKDF  60
TKIFLSFAPD VIINPAAYTA VDKAEDEPHN AFAINTEGAG AVAAAAESIG VPCIHISTDY 120
VFDGSSQTPI CETSPTNPLG IYGKSKLAGE KKVMSCTDNY VILRTAWVYS LFGTNFFLSM 180
LRLSKEYREI SVVCDQFGTP TSASQIAIAV IQIARNLVNN SDNSLRGIFH MTANGGPVSW 240
ADFAEYIFLM STELGGNSCK VRRISTNQYP TKARRPAYSF LDCSKLEKIH NVRISSWKEG 300
VRNVFIK                                                          314
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.

## $His_6$–Serralysin (expressed from pOPINF)

```
MAHHHHHHSS GLEVLFQGPD DIKPVYTTES IADYLLRDSY WANYFLKYQP TVYTDHRLSF  60
DVTELNANGQ EIVRWAFREW SKVVNITFEE TSNLSDIQFK SNDFGYICIP KNNFRYFMEI 120
NLDARDVDFY GIDKGTILSH NAIHEIGHAL GLLHAGPYNG GSPTYGVDNQ YANDSFLTSV 180
MSYFDQNENR TTDASFGYCA TPMVADIVAV QKIYGAPKTL VGDTIHKIEV KQGYSPFIQT 240
IYDSGGSDLL DVQSVGTFDN YIDLNPESWS SIGGYKKNVT IARDTIIESV IGGAGNDTII 300
GNSANNTLFK SGGNDTFYGA DGWDVMVYSM LKNDYKIYRF KNEAIIYDMF RNEVDDIIDV 360
ESLRFSNDQI DLADIKQQSI LEYTASYGDL VQCIGNDVNA SKKHFVDFGL REGRDITFNP 420
YLYLGSYDDL RQVFGSDKDS ATRHYITHGF REGRNPGFMY RDDSMQHSFS LSEKSQEHKY 480
TSTFFDPKED LSENFFEWPN SFQDTIFDFS E                                518
```

Red indicates $His_6$ tag and green indicates HRV 3C protease recognition site, which cleaves between the underlined amino acids.